



(11) **EP 3 067 884 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
08.05.2019 Bulletin 2019/19

(51) Int Cl.:
G10L 15/22^(2006.01) G06F 1/32^(2019.01)

(21) Application number: **16157937.0**

(22) Date of filing: **29.02.2016**

(54) **SPEECH RECOGNITION SYSTEM AND SPEECH RECOGNITION METHOD THEREOF**
SPRACHERKENNUNGSSYSTEM UND SPRACHERKENNUNGSVERFAHREN DAFÜR
PROCÉDÉ DE RECONNAISSANCE VOCALE ET SYSTÈME DE RECONNAISSANCE VOCALE ASSOCIÉ

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **13.03.2015 US 201562132909 P**
29.01.2016 KR 20160011838

(43) Date of publication of application:
14.09.2016 Bulletin 2016/37

(73) Proprietor: **Samsung Electronics Co., Ltd.**
Gyeonggi-do 16677 (KR)

(72) Inventors:
• **KIM, Tae-yoon**
Seoul (KR)

- **KIM, Sang-ha**
Seoul (KR)
- **HAN, Chang-woo**
Seoul (KR)
- **LEE, Jae-won**
Seoul (KR)

(74) Representative: **Appleyard Lees IP LLP**
15 Clare Road
Halifax HX1 2HY (GB)

(56) References cited:
EP-A2- 2 669 889 EP-A2- 2 816 554
US-A1- 2014 281 628 US-B1- 8 543 834

EP 3 067 884 B1

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

1. Field

[0001] Apparatuses and methods consistent with exemplary embodiments relate to speech recognition, and more particularly, to speech recognition based on a wake-up keyword.

2. Description of the Related Art

[0002] A number of the smart devices having speech recognition to enable the functions of the devices to be executed by using a user's speech signal has been steadily increasing.

[0003] In order to enable the speech recognition function of the device, the speech recognition function of the device needs to be activated. A related art speech recognition function is activated by using a fixed wake-up keyword. Accordingly, when a plurality of devices having the same speech recognition function are present in a proximity to each other, a speech recognition function of an unintended device may be activated by the user using the fixed wake-up keyword.

[0004] In addition, a related art speech recognition function separately processes a wake-up keyword and a voice command of a user. Thus, after inputting a wake-up keyword, a user needs to input a voice command after a speech recognition function of a device is activated. If a user continuously or substantially contemporaneously inputs a wake-up keyword and a voice command for the same device or different devices, the related art speech recognition function might not be activated or activated correctly, or a speech recognition error may occur with respect to the input voice command even though the speech recognition function is activated.

[0005] Therefore, there is a need for methods and apparatuses capable of accurately recognizing a user voice command while reliably initiating a speech recognition function of a device.

[0006] EP 2669889 describes a related method and apparatus for executing voice command in an electronic device.

SUMMARY

[0007] The invention is defined in the appended claims. All occurrences of the word "embodiment(s)", except the ones related to the claims, refer to examples useful for understanding the invention which were originally filed but which do not represent embodiments of the presently claimed invention. These examples are shown for illustrative purposes only.

[0008] Exemplary embodiments address at least the above problems and/or disadvantages and other disadvantages not described above. Also, the exemplary embodiments are not required to overcome the disadvantages described above, and may not overcome any of

the problems described above.

[0009] One or more exemplary embodiments provide convenient and accurate speech recognition functions that continuously recognize a personalized wake-up keyword and a voice command.

[0010] One or more exemplary embodiments provide speech recognition functions that are activated more effectively by using a personalized wake-up keyword.

[0011] One or more exemplary embodiments provide speech recognition functions that are activated more effectively by using a personalized wake-up keyword according to device-based environment information.

[0012] According to an aspect of an exemplary embodiment, there is provided a device as set out in accompanying claim 1.

[0013] According to an aspect of an exemplary embodiment, there is provided a speech recognition server as set out in accompanying claim 8.

[0014] According to an aspect of a described embodiment, a speech recognition system includes: a device configured to detect a wake-up keyword from a speech signal of a user; and a speech recognition server configured to set a speech recognition model combined with the wake-up keyword model when a wake-up keyword detection/non-detection signal and the speech signal of the user are received from the device, and recognize the speech signal of the user by using the set speech recognition model, and transmit a speech recognition result to the device.

[0015] According to an aspect of an exemplary embodiment, there is provided a speech recognition method as set out in accompanying claim 10.

[0016] According to an aspect of an exemplary embodiment, there is provided a speech recognition method as set out in accompanying claim 11.

[0017] The wake-up keyword model is one of a plurality of wake-up keyword models which are based on various environment information, the method further comprises receiving, from the device, environment information about the device, and the setting the speech recognition model comprises setting the speech recognition model combined with the wake-up keyword model corresponding to the environment information of the device, among the plurality of wake-up keyword models.

[0018] The method further comprises receiving identification information of the user, from the device, wherein the setting the speech recognition model comprises setting the speech recognition model combined with the wake-up keyword based on the environment information of the device and the identification information of the user.

[0019] According to an aspect of a described embodiment, a speech recognition method, which is performed by a speech recognition system, includes: registering a wake-up keyword model in a device and a speech recognition server; when a speech signal of a user is received through the device, detecting the wake-up keyword from the speech signal of the user by using the wake-up keyword model; transmitting a wake-up key-

word detection/non-detection signal and the speech signal of the user from the device to the speech recognition server; setting, by the speech recognition server, a speech recognition model according to the wake-up keyword detection/non-detection signal; recognizing, by the speech recognition server, the speech signal of the user by using the set speech recognition model; removing, by the speech recognition server, the wake-up keyword from a result of recognizing the speech signal of the user; transmitting the result of recognizing the speech signal of the user, from which the wake-up keyword is removed, from the speech recognition server to the device; and controlling, by the device, the device according to the received result of recognizing the speech signal of the user.

[0020] According to an aspect of a described embodiment, a device comprising: an audio input receiver configured to receive an audio signal from a user, the audio signal including a wake-up keyword; a memory configured to store wake-up keyword models for identifying the wake-up keyword from the received audio signal; and a processor configured to perform: detecting the wake-up keyword from the received audio signal by matching the wake-up keyword included into the received audio signal with the stored wake-up keyword models, generating a detection value indicative whether the wake-up keyword has been detected or not, based on a result of the matching, transmitting the detection value and the received audio signal to a server, receiving, from the server, a speech recognition result of the audio signal that is translated based on the detection value, and controlling an executable application of the device in performing a device function, based on the speech recognition result.

[0021] The detection value indicates that the wake-up keyword has been detected in the received audio signal, and the processor is configured to receive the speech recognition result which includes a user command to execute the application, and in which the wake-up keyword itself is absent.

[0022] The audio input receiver is configured to receive, in advance, individual user inputs containing individual keywords related to a control of executable applications of the device, and the memory is configured to store the wake-up keyword models based on the received individual keywords.

[0023] According to an aspect of a described embodiment, a method comprises: storing, in a first memory, wake-up keyword models which are used to identify wake-up keywords; receiving an audio signal from a user, the audio signal including the wake-up keyword; detecting the wake-up keyword from the received audio signal by matching the wake-up keyword included into the received audio signal with the stored wake-up keyword models; generating a detection value indicative whether the wake-up keyword has been detected or not based on a result of the matching; transmitting the detection value and the received audio signal to a server; receiving, from the server, a speech recognition result of the audio

signal that is translated based on the detection value; and controlling an executable application of the device in performing a device function, based on the speech recognition result.

[0024] The method further comprises storing, in a second memory, a speech recognition model which is used to translate the audio signal of the user and the wake-up keyword models which are synchronized with the wake-up keyword models stored in the first memory, wherein the receiving the speech recognition result comprises: identifying, by the server, whether the audio signal contains the wake-up keyword, from the detection value; and translating, by the server, the audio signal into the speech recognition result based on a combination model in which the speech recognition model is combined with a respective wake-up keyword model in response to the detection value indicating that the audio signal contains the wake-up keyword.

[0025] The receiving the speech recognition result further comprises: generating, by the server, the speech recognition result by removing the wake-up keyword from the speech recognition result, and receiving, from the server, the speech recognition result of the audio signal from which the wake-up keyword has been removed; wherein the controlling comprises controlling the executable application of the device according to the speech recognition result from which the wake-up keyword has been removed.

[0026] The translating comprises translating the audio signal into the speech recognition result by using the speech recognition model only, in response to the detection value indicating that the audio signal does not contain the wake-up keyword.

35 BRIEF DESCRIPTION OF THE DRAWINGS

[0027] The above and/or other aspects will become more apparent by describing certain exemplary embodiments with reference to the accompanying drawings, in which:

FIG. 1 is a diagram for describing a speech recognition system according to an exemplary embodiment;

FIG. 2 is a flowchart of a speech recognition method according to an exemplary embodiment, which is performed based on a device and a speech recognition server included in the speech recognition system;

FIG. 3 is a flowchart of a process of registering a wake-up keyword model in the speech recognition method according to an exemplary embodiment;

FIG. 4 is a flowchart of another process of registering a wake-up keyword model in the speech recognition method according to an exemplary embodiment;

FIGS. 5A and 5B illustrate examples of candidate wake-up keyword models displayed on a display of the device included in the speech recognition system

according to an exemplary embodiment;
 FIGS. 6 and 7 are flowcharts of speech recognition methods according to an exemplary embodiment, which are performed based on the device and the speech recognition server included in the speech recognition system;
 FIG. 8 is a flowchart of a speech recognition method according to an exemplary embodiment, which is performed by a device;
 FIGS. 9 and 10 are configuration diagrams of the device included in the speech recognition system according to an exemplary embodiment;
 FIG. 11 is a configuration diagram of the speech recognition server included in the speech recognition system according to an exemplary embodiment; and
 FIG. 12 is a configuration diagram of a speech recognition system according to an exemplary embodiment.

DETAILED DESCRIPTION

[0028] Certain exemplary embodiments are described in greater detail below with reference to the accompanying drawings.

[0029] In the following description, like drawing reference numerals are used for like elements, even in different drawings. The matters defined in the description, such as detailed construction and elements, are provided to assist in a comprehensive understanding of the exemplary embodiments. However, it is apparent that the exemplary embodiments can be practiced without those specifically defined matters. Also, well-known functions or constructions are not described in detail since they would obscure the description with unnecessary detail.

[0030] As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed items.

[0031] It will be understood that when a region is referred to as being "connected to" or "coupled to" another region, it may be directly connected or coupled to the other region or intervening regions may be present. It will be understood that terms such as "comprise", "include", and "have", when used herein, specify the presence of stated elements, but do not preclude the presence or addition of one or more other elements.

[0032] The term "wake-up keyword" used herein refers to information capable of activating or initiating a speech recognition function. The wake-up keyword used herein may be referred to as a wake-up word. The wake-up keyword used herein may be based on a speech signal of a user, but is not limited thereto. For example, the wake-up keyword used herein may include a sound (or audio signal) based on a gesture of a user.

[0033] The sound based on the gesture of the user may include, for example, a sound generated when the user strikes his/her fingers together. The sound based on the gesture of the user may include, for example, a sound generated when the user clicks his/her tongue.

The sound based on the gesture of the user may include, for example, a sound of a user's laughter. The sound based on the gesture of the user may include, for example, a sound generated when the user's lips quiver. The sound based on the gesture of the user may include, for example, a sound of a user's whistle. The sound based on the gesture of the user is not limited to those described above.

[0034] When the wake-up keyword used herein includes the sound based on the gesture of the user, the wake-up keyword may be referred to as a wake-up keyword signal.

[0035] A wake-up keyword model used herein may refer to a wake-up keyword preregistered in a device and/or a speech recognition server so as to detect or recognize a wake-up keyword. The wake-up keyword model may include a personalized acoustic model and/or a personalized language model, but is not limited thereto. The acoustic model models signal characteristics of the speech of the user (or sound based on the gesture of the user). The language model models a linguistic sequence of words or syllables corresponding to recognition vocabulary.

[0036] Since the wake-up keyword model registered in the device is used for detecting the wake-up keyword, the wake-up keyword model used herein may be referred to as a model for wake-up keyword detection. Since the wake-up keyword model registered in the speech recognition server is used for detecting the wake-up keyword, the wake-up keyword model used herein may be referred to as a model for wake-up keyword recognition.

[0037] The model for wake-up keyword detection and the model for wake-up keyword recognition may be identical to or different from each other. For example, when the model for wake-up keyword detection includes an acoustic model corresponding to a personalized wake-up keyword "Hi", the model for wake-up keyword recognition may include, for example, an acoustic model corresponding to the personalized wake-up keyword "Hi" and a tag, e.g., "!", which is associated with and identifies the wake-up keyword. The model for wake-up keyword detection and the model for wake-up keyword recognition are not limited to those described above.

[0038] The model for wake-up keyword detection and the model for wake-up keyword recognition may be referred to as the wake-up keyword model. However, the wake-up keyword model registered in the device may be understood as the model for wake-up keyword detection, and the wake-up keyword model registered in the speech recognition server may be understood as the model for wake-up keyword recognition.

[0039] The wake-up keyword model may be generated by the device or the speech recognition server. The device or the speech recognition server may transmit and receive data so as to share the generated wake-up keyword model with each other.

[0040] The speech recognition function used herein may refer to converting the speech signal of the user into

a character string or a text. The text may be human-perceptible phrase, sentence, or a set of words. The speech signal of the user may include a voice command. The voice command may execute a particular function of the device.

[0041] The particular function of the device used herein may include executing an executable application set in the device, but is not limited thereto.

[0042] For example, when the device is a smart phone, the executing of the application may include a phone call, route finding, Internet browsing, alarm setting, and/or any other appropriate executable function available in the smart phone. When the device is a smart television (TV), the executing of the application may include program searching, channel searching, Internet browsing, and/or any other appropriate executable function available in the smart TV. When the device is a smart oven, the executing of the application may include recipe searching, etc. When the device is a smart refrigerator, the executing of the application may include chilling state checking, freezing state checking, etc. When the device is a smart vehicle, the executing of the application may include automatic start-up, autonomous cruise, automatic parking, automatic media device turn on and turn off, automatic air control, etc. The described-above examples of the executable applications are not limited thereto.

[0043] The voice command used herein may be a word, a sentence, or a phrase. The speech recognition model used herein may include a personalized acoustic model and/or a personalized language model.

[0044] FIG. 1 is a diagram for describing a speech recognition system 10 according to an exemplary embodiment. The speech recognition system 10 may include a device 100 and a speech recognition server 110.

[0045] The device 100 may receive a speech signal from a user 101 that may include the wake-up keyword and the voice command. The device 100 may detect a wake-up keyword from the received speech signal of the user 101 by using a wake-up keyword model. The device 100 may generate, in advance, the wake-up keyword model, and register and store the generated wake-up keyword model in the device 100. The device 100 may transmit the generated wake-up keyword model to the speech recognition server 110. As another example, the device 100 may store the wake-up keyword model which has been received from the speech recognition server 110.

[0046] The device 100 may use the registered wake-up keyword model or the wake-up keyword model to detect the wake-up keyword from the received speech signal of the user 101. However, the speech signal of the user 101, which is received by the device 100, might not include the wake-up keyword or the device 100 might not be able to match the wake-up keyword with the stored wake-up keyword model.

[0047] The device 100 may generate and transmit a wake-up keyword detection/non-detection signal and the received speech signal of the user 101 to the speech

recognition server 110. The wake-up keyword detection/non-detection signal is a signal that indicates whether the wake-up keyword has been detected from the received speech signal of the user 101.

[0048] The device 100 may represent the wake-up keyword detection/non-detection signal by binary data. When the wake-up keyword has been detected from the received speech signal of the user 101, the device 100 may represent the wake-up keyword detection/non-detection signal by, for example, "0". When the wake-up keyword has not been detected from the received speech signal of the user 101, the device 100 may represent the wake-up keyword detection/non-detection signal by, for example, "1".

[0049] The speech recognition server 110 may receive the wake-up keyword detection/non-detection signal and the speech signal of the user 101 from the device 100. The speech signal of the user 101, which is received from the device 100, may be substantially the same as the speech signal of the user 101, which is received by the device 100. Additionally, the device 100 may transmit the wake-up keyword model.

[0050] The speech recognition server 110 may set the speech recognition model according to the received wake-up keyword detection/non-detection signal. When the wake-up keyword detection/non-detection signal indicates that the wake-up keyword is included in the speech signal of the user 101, the speech recognition server 110 may set the speech recognition model to recognize the speech signal of the user 101 by using a combination model in which the speech recognition model is combined with the wake-up keyword model stored or received by the server 110.

[0051] In the speech recognition server 110, the wake-up keyword model that is combined with the speech recognition model is matched with the wake-up keyword detected by the device 100. For example, when the wake-up keyword detected by the device 100 is "Hi", the speech recognition server 110 may set the speech recognition model to recognize the speech signal of the user 101 by using "Hi + speech recognition model (for example, play the music)". When the wake-up keyword model is combined with the speech recognition model, the speech recognition server 110 may take into consideration a silence duration between the wake-up keyword model and the speech recognition model.

[0052] As described above, the speech recognition server 110 may stably secure the speech signal of the user 101 by continuously performing a recognition process on the wake-up keyword and the voice command included in the speech signal of the user 101, thereby improving the speech recognition performance of the speech recognition system 10.

[0053] When the wake-up keyword detection/non-detection signal indicates that the wake-up keyword is not included in the speech signal of the user 101, the speech recognition server 110 may set the speech recognition model to recognize the speech signal of the user 101 by

using the speech recognition model that is not combined with the wake-up keyword model. Optionally, the speech recognition server 110 may verify that the wake-up keyword is included or is not included in the received speech signal of the user 101.

[0054] The speech recognition server 110 may dynamically reconfigure (or switch) the speech recognition model that is used for recognizing the speech of the user 101 according to the wake-up keyword detection/non-detection signal. Therefore, the setting of the speech recognition model according to the wake-up keyword detection/non-detection signal, which is performed by the speech recognition server 110, may be determining the configuration of the speech recognition model according to the detection or non-detection of the wake-up keyword.

[0055] The setting of the speech recognition model in the speech recognition server 110 may include loading the speech recognition model. Accordingly, the wake-up keyword detection/non-detection signal may be construed as including a speech recognition model loading request signal, a speech signal model setting request signal, or a speech recognition model loading trigger signal. The expression of the wake-up keyword detection/non-detection signal used herein is not limited to those described above.

[0056] The speech recognition server 110 may generate the speech recognition model for recognizing the voice command. The speech recognition model may include an acoustic model and a language model. The acoustic model models signal characteristics of the speech. The language model models a linguistic sequence relationship of words or syllables corresponding to recognition vocabulary.

[0057] The speech recognition server 110 may detect only a speech part from the received speech signal of the user 101. The speech recognition server 110 may extract speech characteristics from the detected speech part. The speech recognition server 110 may perform a speech recognition process on the received speech signal of the user 101 by using the extracted speech characteristics, characteristics of a preregistered acoustic model, and the language model. The speech recognition server 110 may perform the speech recognition process by comparing the extracted speech characteristics with the characteristics of the preregistered acoustic model. The speech recognition process, which is performed on the received speech signal of the user 101 by the speech recognition server 110, is not limited to those described above.

[0058] The speech recognition server 110 may remove the wake-up keyword from a speech recognition result of the speech recognition process. The speech recognition server 110 may transmit the speech recognition result, from which the wake-up keyword is removed, to the device 100.

[0059] The speech recognition server 110 may generate the wake-up keyword model. The speech recognition server 110 may transmit the generated wake-up keyword

model to the device 100 while registering (or storing) the generated wake-keyword model in the speech recognition server 110. Accordingly, the device 100 and the speech recognition server 110 may share the wake-up keyword model with each other.

[0060] The device 100 may control the function of the device 100 according to the speech recognition result received from the speech recognition server 110.

[0061] When the device 100 or the speech recognition server 110 generates a plurality of wake-up keyword models, the device 100 or the speech recognition server 110 may allocate identification information to each of the wake-up keyword models. When the identification information is allocated to each of the wake-up keyword models, the wake-up keyword detection/non-detection signal transmitted from the device 100 to the speech recognition server 110 may include identification information about the detected wake-up keyword.

[0062] When the device 100 is a portable device, the device 100 may include at least one device among a smart phone, a notebook computer, a smart board, a tablet personal computer (PC), a handheld device, a handheld computer, a media player, an e-book device, and a personal digital assistant (PDA), but is not limited thereto.

[0063] When the device 100 is a wearable device, the device 100 may include at least one device among smart glasses, smart watches, smart bands (e.g., smart waistbands, smart headbands, etc.), various smart accessories (e.g., smart rings, smart bracelets, smart ankle bracelets, smart hairpins, smart clips, smart necklaces, etc.), various smart body protectors (e.g., smart knee pads, smart elbow pads, etc.), smart shoes, smart gloves, smart clothes, smart hats, smart artificial legs, and smart artificial hands, but is not limited thereto.

[0064] The device 100 may include a device based on a Machine-to-Machine (M2M) or Internet of Things (IoT) network (e.g., smart household appliances, smart sensors, etc.), vehicles, and vehicle navigation devices, but is not limited thereto.

[0065] The device 100 and the speech recognition server 110 may be connected to each other via a wired and/or wireless network. The device 100 and the speech recognition server 110 may be connected to each other via a short-range wireless network and/or a long-range wireless network.

[0066] FIG. 2 is a flowchart of a speech recognition method according to an exemplary embodiment, which is performed based on the device 100 and the voice recognition server 110 included in the speech recognition system 10. FIG. 2 illustrates a case where the speech recognition is performed based on the speech signal of the user 101.

[0067] Referring to FIG. 2, in operation S201, the device 100 may register the wake-up keyword model, as described in detail below with reference to FIGS 3 and 4.

[0068] FIG. 3 is a flowchart of registering the wake-up keyword model in the speech recognition method accord-

ing to an exemplary embodiment.

[0069] Referring to FIG. 3, in operation S301, the device 100 may receive the speech signal of the user 101. The speech signal of the user 101, which is received in operation S301, is used for registering the wake-up keyword model. In operation S301, the device 100 may receive a sound (or audio signal) based on a particular gesture of the user 101, instead of the speech signal of the user 101.

[0070] In operation S302, the device 100 may recognize the speech signal of the user 101 by using the speech recognition model. The speech recognition model may include an acoustic model and/or a language model based on automatic speech recognition (ASR), but is not limited thereto.

[0071] In operation S303, the device 100 may determine whether the received speech signal of the user 101 is valid as the wake-up keyword model, based a speech matching rate of the speech signal of the user 101.

[0072] For example, in a case where the device 100 recognizes the speech signal of the user 101 twice or more times and compares recognition results, if consistent results come out a preset number of times or more, the device 100 may determine that the received speech signal of the user 101 is valid as the wake-up keyword model.

[0073] When it is determined in operation S303 that the received speech signal of the user 101 is valid as the wake-up keyword model, the device 100 may generate and/or register the wake-up keyword model in the device 100 in operation S304. The registering of the wake-up keyword model in the device 100 may mean storing the wake-up keyword model in the device 100.

[0074] In operation S303, in a case where the device 100 recognizes the speech signal of the user 101 twice or more times and compares recognition results, if a number of consistent recognition results is below the preset number of times, the device 100 may determine that the received speech signal of the user 101 is invalid as the wake-up keyword model.

[0075] When it is determined in operation S303 that the received speech signal of the user 101 is invalid as the wake-up keyword model, the device 100 does not register the received speech signal of the user 101 as the wake-up keyword model.

[0076] When it is determined in operation S303 that the received speech signal of the user 101 is invalid as the wake-up keyword model, the device 100 may output a notification message. The notification message may have various forms and contents. For example, the notification message may include a message indicating that "the currently input speech signal of the user 101 is not registered as the wake-up keyword model." The notification message may include a message guiding for the user 101 to input a speech signal that is registrable as the wake-up keyword model.

[0077] FIG. 4 is a flowchart of registering the wake-up keyword model in the speech recognition method accord-

ing to an exemplary embodiment.

[0078] In operation S401, the device 100 may request candidate wake-up keyword models stored therein. The request for the candidate wake-up keyword models may be based on the speech signal of the user 101, but is not limited thereto. For example, the device 100 may receive a user input of requesting the candidate wake-up keyword models according to a particular button control (or dedicated button) of the device 100 or a touch-based input.

[0079] In operation S402, the device 100 may output the candidate wake-up keyword models. The device 100 may output the candidate wake-up keyword models through a display of the device 100.

[0080] FIGS. 5A and 5B illustrate examples of candidate wake-up keyword models displayed on the display of the device 100 included in the speech recognition system 10 according to an exemplary embodiment.

[0081] FIG. 5A illustrates an example of a candidate wake-up keyword model list displayed on the display 98 of the device 100. Referring to FIG. 5A, the candidate wake-up keyword model is provided in a text form.

[0082] When a touch-based input of selecting a first candidate wake-up keyword model is received based on the candidate wake-up keyword model list illustrated in FIG. 5A, the device 100 may output an audio signal corresponding to the selected first candidate wake-up keyword model while displaying a speech waveform of the selected first candidate wake-up keyword model as illustrated in FIG. 5B. Accordingly, before selecting the wake-up keyword model, the user 101 may confirm the wake-up keyword model to be selected.

[0083] In operation S402, the device 100 may output the candidate wake-up keyword models through an audio output transmitter (e.g., speaker) of the device 100.

[0084] When a selection signal of selecting one candidate wake-up keyword model among the candidate wake-up keyword models is received in operation S403, the device 100 may automatically generate and/or register the selected candidate wake-up keyword model in operation S404. As another example, the device 100 may request the input of the speech signal of the user 101 corresponding to the selected candidate wake-up keyword model, generate the received speech signal of the user 101 as the wake-up keyword model, and/or register the wake-up keyword model.

[0085] Referring again to FIG. 2, in operation S201, the device 100 may set a communication channel for the speech recognition server 110 and request the wake-up keyword model while transmitting the received speech signal of the user 101 to the speech recognition server 110 via the set communication channel. Accordingly, the device 100 may receive the wake-up keyword model generated by the speech recognition server 110.

[0086] In operation S202, the speech recognition server 110 may register the wake-up keyword model. In operation S202, the speech recognition server 110 may register the wake-up keyword model received from the de-

vice 100, but the method of registering the wake-up keyword model in the speech recognition server 110 is not limited to those described above.

[0087] For example, the speech recognition server 110 may request the device 100 to send the wake-up keyword model and receive the wake-up keyword model. To this end, the speech recognition server 110 may monitor the device 100. The speech recognition server 110 may periodically monitor the device 100. When the approach of the device 100 is detected, the speech recognition server 110 may monitor the device 100. When the connection of the device 100 to the speech recognition server 110 is detected, the speech recognition server 110 may request the device 100 to send the wake-up keyword model.

[0088] In operation S202, when the wake-up keyword model is registered, the speech recognition server 110 may add a tag, which identifies the wake-up keyword, to the wake-up keyword model. The tag may be represented by a special symbol (e.g., !), but is not limited thereto.

[0089] In operation S202, the wake-up keyword model registered in the speech recognition server 110 may be synchronized with the wake-up keyword model registered in the device 100. When the wake-up keyword model registered in the device 100 is updated, the wake-up keyword model registered in the speech recognition server 110 may be updated.

[0090] As another example, in operation S202, the speech recognition server 110 may receive the speech signal of the user 101 from the device 100 and generate and register the wake-up keyword model, prior to operation S201. The speech recognition server 110 may generate the wake-up keyword model, as described above with reference to FIG. 3 or 4.

[0091] In operation S203, the device 100 may receive the speech signal of the user 101. In operation S204, the device 100 may detect the wake-up keyword from the received speech signal of the user 101 by using the registered wake-up keyword model. The device 100 may detect the wake-up keyword by comparing signal characteristics between the registered wake-up keyword model and the received speech signal of the user 101.

[0092] In operation S205, the device 100 may transmit the wake-up keyword detection/non-detection signal and the received speech signal of the user 101 to the speech recognition server 110.

[0093] In operation S206, the speech recognition server 110 may set the speech recognition model according to the received wake-up keyword detection/non-detection signal. The setting of the speech recognition model may be the same as described with reference to FIG. 1. That is, when the wake-up keyword detection/non-detection signal indicates that the wake-up keyword has been detected, the speech recognition server 110 may set the speech recognition model that is combined with the wake-up keyword model. When the wake-up keyword detection/non-detection signal indicates that the wake-up keyword has not been detected, the speech recogni-

tion server 110 may set the speech recognition model that is not combined with the wake-up keyword model.

[0094] In operation S207, the speech recognition server 110 may recognize the received speech signal of the user 101 by using the set speech recognition model. In operation S208, the speech recognition server 110 may remove the wake-up keyword from the speech recognition result. The speech recognition server 110 may remove the wake-up keyword from the speech recognition result by using the tag added to the wake-up keyword when the wake-up keyword model is registered.

[0095] In operation S209, the speech recognition server 110 may transmit the speech recognition result, from which the wake-up keyword is removed, to the device 100. In operation S210, the device 100 may control the device 100 according to the received speech recognition result.

[0096] FIG. 6 is a flowchart of a speech recognition method according to an exemplary embodiment, which is performed based on the device 100 and the speech recognition server 110 included in the speech recognition system 10. FIG. 6 illustrates an example of speech recognition that is performed by using a wake-up keyword model according to environment information based on the device 100.

[0097] In operation S601, the device 100 may register a plurality of wake-up keyword models based on environment information. The environment information may include location information. The location information may include physical location information and logical location information. The physical location information refers to information represented by latitude and longitude. The logical location information refers to information represented by semantic information such as home, office, or cafe. The environment information may include weather information. The environment information may include time information. The environment information may include schedule information. The environment information may include location, time, weather, and/or schedule information. The environment information is not limited thereto, and the environment information may include condition information or situation information that directly or indirectly influence the user 101.

[0098] For example, the device 100 may differently register the wake-up keyword model when the location of the device 100 is home and the wake-up keyword model when the location of the device 100 is office. The device 100 may differently register the wake-up keyword model when the time detected by the device 100 is 6 a.m. and the wake-up keyword model when the time detected by the device 100 is 6 p.m. The device 100 may differently register the wake-up keyword model when the weather detected by the device 100 is fine and the wake-up keyword model when the weather detected by the device 100 is rainy. The device 100 may register a different wake-up keyword model according to the schedule of the user 101 that is detected by the device 100.

[0099] In operation S601, the device 100 may receive

the plurality of wake-up keyword models from the speech recognition server 110, based on the environment information and register the plurality of wake-up keyword models, as described in operation S201.

[0100] In operation S602, the speech recognition server 110 may register a plurality of wake-up keyword models based on environment information.

[0101] The plurality of wake-up keyword models registered in the speech recognition server 110 may be synchronized in real time with the plurality of wake-up keyword models registered in the device 100. Accordingly, whenever the plurality of wake-up keyword models registered in the device 100 are updated, the plurality of wake-up keyword models registered in the speech recognition server 110 may be updated.

[0102] In operation S602, the speech recognition server 110 may register the plurality of wake-up keyword models received from the device 100. In operation S602, the speech recognition server 110 may request the device 100 to send the plurality of wake-up keyword models and receive the plurality of wake-up keyword models from the device 100.

[0103] In operation S602, as described in operation S202, the speech recognition server 110 may set a communication channel between the device 100 and the speech recognition server 110 and generate and register the plurality of wake-up keyword models based on the above-described environment information by using the speech signal of the user 101 received from the device 100 via the set communication channel. The speech recognition server 110 may provide the device 100 with the plurality of registered wake-up keyword models.

[0104] In operation S603, the device 100 may receive the speech signal of the user 101. In operation S604, the device 100 may detect environment information based on the device 100. The device 100 may detect the environment information based on the device 100 by using sensors included in the device 100 or applications set in the device 100.

[0105] For example, the device 100 may detect location information by using a position sensor (e.g., global positioning system (GPS) sensor) included in the device 100. The device 100 may detect time information by using a timer application set in the device 100. The device 100 may detect weather information by using a weather application set in the device 100. The device 100 may detect a schedule of the user 101 by using a schedule application set in the device 100.

[0106] In operation S605, the device 100 may detect the wake-up keyword from the received speech signal of the user 101 by using the wake-up keyword model corresponding to the detected environment information among the plurality of registered wake-up keyword models.

[0107] For example, in a case where the wake-up keyword model at the home is "Hi" and the wake-up keyword model in the office is "Good", if the location of the device 100 detected by the device 100 is office, the device 100

may detect the wake-up keyword from the received speech signal of the user 101 by using "Good".

[0108] In operation S606, the device 100 may transmit the detected environment information, the wake-up keyword detection/non-detection signal, and the received speech signal of the user 101 to the speech recognition server 110.

[0109] In operation S607, the speech recognition server 110 may determine the wake-up keyword model according to the wake-up keyword detection/non-detection signal and the received environment information based on the device 100, and set the speech recognition model that is combined with the determined wake-up keyword model.

[0110] In operation S608, the speech recognition server 110 may recognize the received speech signal of the user 101 by using the set speech recognition model. In operation S609, the speech recognition server 110 may remove the wake-up keyword from the speech recognition result. The speech recognition server 110 may remove the wake-up keyword from the speech recognition result by using the tag added to the wake-up keyword when the wake-up keyword model is registered.

[0111] In operation S610, the speech recognition server 110 may transmit the speech recognition result, from which the wake-up keyword is removed, to the device 100. In operation S611, the device 100 may control the device 100 according to the received speech recognition result.

[0112] FIG. 7 is a flowchart of a speech recognition method according to an exemplary embodiment, which is performed based on the device 100 and the speech recognition server 110 included in the speech recognition system 10. FIG. 7 illustrates an example of speech recognition that is performed by setting the speech recognition model according to the identification information of the user 101, the environment information based on the device 100, and the wake-up keyword detection/non-detection signal.

[0113] In operation S701, the device 100 may register a plurality of wake-up keyword models based on environment information. The environment information may be substantially the same as described in operation S601 of FIG. 6, but is not limited thereto. In operation S701, the device 100 may register the plurality of wake-up keyword models received from the speech recognition server 110.

[0114] In operation S702, the speech recognition server 110 may register a plurality of wake-up keyword models based on the environment information and the identification information of the user 101. For example, the speech recognition server 110 may register a plurality of wake-up keyword models based on environment information with respect to identification information A of the user 101. The speech recognition server 110 may register a plurality of wake-up keyword models based on environment information with respect to identification information B of the user 101.

[0115] The plurality of wake-up keyword models registered in the speech recognition server 110 may be synchronized with respect to each user. For example, when the plurality of wake-up keyword models of a user A are updated, the plurality of wake-up keyword models of the user A among the plurality of wake-up keyword models registered in the speech recognition server 110 are also updated.

[0116] In operation S702, the speech recognition server 110 may register the wake-up keyword models based on the speech signal of the user 101 received from the device 100. In this case, the speech recognition server 110 may provide the device 100 with the plurality of registered wake-up keyword models.

[0117] In operation S703, the device 100 may receive the speech signal of the user 101. In operation S704, the device 100 may detect the environment information based on the device 100. In operation S705, the device 100 may acquire the identification information of the user 101 based on the received speech signal of the user 101. The identification information of the user 101 may include a nickname, a gender, and a name of the user 101, but is not limited thereto.

[0118] In operation S705, the identification information of the user 101 may be acquired by using a fingerprint recognition technique or an iris recognition technique.

[0119] In operation S706, the device 100 may detect the wake-up keyword from the received speech signal of the user 101 by using the wake-up keyword model corresponding to the detected environment information among the plurality of registered wake-up keyword models.

[0120] In operation S707, the device 100 may transmit the detected environment information, the identification information of the user 101, the wake-up keyword detection/non-detection signal, and the received speech signal of the user 101 to the speech recognition server 110.

[0121] In operation S708, the speech recognition server 110 may determine the wake-up keyword model according to the wake-up keyword detection/non-detection signal, the received environment information based on the device 100, and the identification information of the user 101, and set the speech recognition model that is combined with the determined wake-up keyword model.

[0122] In operation S709, the speech recognition server 110 may recognize the received speech signal of the user 101 by using the set speech recognition model. In operation S710, the speech recognition server 110 may remove the wake-up keyword from the speech recognition result. The speech recognition server 110 may remove the wake-up keyword from the speech recognition result by using the tag added to the wake-up keyword when the wake-up keyword model is registered.

[0123] In operation S711, the speech recognition server 110 may transmit the speech recognition result, from which the wake-up keyword is removed, to the device 100. In operation S712, the device 100 may control the device 100 according to the received speech recognition

result.

[0124] FIG. 8 is a flowchart of a speech recognition method according to an exemplary embodiment, which is performed by the device 100. FIG. 8 illustrates a case where the speech recognition is performed by the device 100, regardless of the speech recognition server 110.

[0125] In operation S801, the device 100 may register a wake-up keyword model. When the wake-up keyword model is registered, the device 100 may add a tag to a wake-up keyword so as to identify the wake-up keyword. In operation S801, the device 100 may receive the wake-up keyword model from the speech recognition server 110 and register the received wake-up keyword model.

[0126] In operation S802, the device 100 may receive a speech signal of a user 101. In operation S803, the device 100 may detect the wake-up keyword from the speech signal of the user 101 by using the wake-up keyword model.

[0127] When it is determined in operation S804 that the wake-up keyword is detected, the device 100 proceeds to operation S805 to set the speech recognition model that is combined with the wake-up keyword model. In operation S806, the device 100 may perform a speech recognition process on the received speech signal of the user 101 by using the speech recognition model.

[0128] In operation S807, the device 100 may remove the wake-up keyword from the speech recognition result. The device 100 may remove the wake-up keyword from the speech recognition result by using the tag that identifies the wake-up keyword. In operation S808, the device 100 may control the device 100 according to the speech recognition result, from which the wake-up keyword is removed.

[0129] When it is determined in operation S804 that the wake-up keyword is not detected, the device 100 proceeds to operation S809 to set the speech recognition model that is not combined with the wake-up keyword model. In operation S810, the device 100 may perform a speech recognition process on the received speech signal of the user 101 by using the speech recognition model. In operation S811, the device 100 may control the device 100 according to the speech recognition result.

[0130] The speech recognition method of FIG. 8 may be modified to register a plurality of wake-up keyword models based on environment information and recognize a speech signal, as described with reference to FIG. 6.

[0131] The speech recognition method of FIG. 2, 6, 7, and/or 8 may be modified to register a plurality of wake-up keyword models, regardless of environment information, and recognize a speech signal. The plurality of wake-up keyword models may be set for each user. When the plurality of wake-up keyword models are registered, each of the wake-up keyword models may include identification information capable of identifying the wake-up keyword.

[0132] FIG. 9 is a functional block diagram of a device 100 according to an exemplary embodiment.

[0133] Referring to FIG. 9, the device 100 may include an audio input receiver 910, a communicator 920, a processor 930, a display 940, a user input receiver 950, and a memory 960.

[0134] The audio input receiver 910 may receive a speech signal of a user 101. The audio input receiver 910 may receive a sound (audio signal) based on a particular gesture of the user 101.

[0135] The audio input receiver 910 may receive an audio signal that is input from the outside of the device 100. The audio input receiver 910 may convert the received audio signal into an electrical audio signal and transmit the electrical audio signal to the processor 930. The audio input receiver 910 may be configured to perform operations based on various noise removal algorithms for removing noise generated in the process of receiving external acoustic signals. The audio input receiver 910 may include a microphone.

[0136] The communicator 920 may be configured to connect the device 100 to the speech recognition server 110 via a wired and/or wireless network. The communicator 920 may be implemented to have substantially the same configuration as a communicator 1040 to be described with reference to FIG. 10.

[0137] The processor 930 may be a controller that controls the operation of the device 100. The processor 930 may control the audio input receiver 910, the communicator 920, the display 940, the user input receiver 950, and the memory 960. When the speech signal of the user 101 is received through the audio input receiver 910, the processor 930 may perform a speech recognition process using the wake-up keyword model in real time.

[0138] The processor 930 may register the wake-up keyword model in the memory 960. The processor 930 may register, in the memory 960, the wake-up keyword model received from the speech recognition server 110 via the communicator 920. The processor 930 may request the wake-up keyword model based on the speech signal of the user 101 while transmitting the speech signal of the user, which is received via the audio input receiver 910, to the speech recognition server 110.

[0139] The processor 930 may transmit the wake-up keyword model registered in the memory 960 to the speech recognition server 110 via the communicator 920. When a wake-up keyword model request signal is received from the speech recognition server 110 via the communicator 920, the processor 930 may transmit the registered wake-up keyword model to the speech recognition server 110. The processor 930 may transmit the registered wake-up keyword model to the speech recognition server 110 at the same time as the registration of the wake-up keyword model in the memory 960.

[0140] The processor 930 may detect the wake-up keyword from the received speech signal of the user 101 by using the wake-up keyword model registered in the memory 960 as the speech signal of the user 101 is received through the audio input receiver 910. The processor 930 may transmit a wake-up keyword detection/non-detection

signal and the received speech signal of the user 101 to the speech recognition server 110 via the communicator 920.

[0141] The processor 930 may receive the speech recognition result from the speech recognition server 110 via the communicator 920. The processor 930 may control the device 100 according to the received speech recognition result.

[0142] When an audio signal for registering the wake-up keyword model is received through the audio input receiver 910, the processor 930 may determine whether the audio signal is usable as the wake-up keyword model, based on a matching rate of the audio signal, as described above.

[0143] The processor 930 may register a candidate wake-up keyword model, which is selected among candidate wake-up keyword models stored in the memory 960, in the memory 960 according to a user input received through the user input receiver 950.

[0144] The processor 930 may include a main processor and a sub processor according to an implementation type of the device 100. The sub processor may be set as a low-power processor.

[0145] The display 940 may be configured to display the candidate wake-up keyword requested by the user 101 under the control of the processor 930. The display 940 may include a liquid crystal display (LCD), a thin-film transistor-liquid crystal display (TFT-LCD), an organic light-emitting diode (OLED), a flexible display, a three-dimensional (3D) display, or an electrophoretic display (EPD). The display 940 may include, for example, a touch screen, but is not limited thereto.

[0146] The user input receiver 950 may be configured to receive a user input with respect to the device 100.

The user input receiver 950 may receive a user input of requesting registration of the wake-up keyword, a user input of selecting one among candidate wake-up keywords, and/or a user input of registering the selected candidate wake-up keyword. The user input received through the user input receiver 950 is not limited thereto. The user input receiver 950 may transmit the received user input to the processor 930.

[0147] The memory 960 may store the wake-up keyword model. The memory 960 may store programs for processing and control of the processor 930. The programs stored in the memory 960 may include an operating system (OS) and various application programs. The various application programs may include the speech recognition program and the camera application. The memory 960 may store information managed by the application programs, for example, wake-up keyword use history information of the user 101, schedule information of the user 101, and/or the profile information of the user 101.

[0148] The programs stored in the memory 960 may include a plurality of modules according to functions thereof. The plurality of modules may include, for example, a mobile communication module, a wireless fidelity

(Wi-Fi) module, a Bluetooth module, a digital multimedia broadcasting (DMB) module, a camera module, a sensor module, a GPS module, a video reproduction module, an audio reproduction module, a power module, a touch screen module, a user interface (UI) module, and/or an application module.

[0149] The memory 960 may include a flash memory, a hard disk, a multimedia card micro type memory, a card type memory (e.g., SD or XD memory), random access memory (RAM), static random access memory (SRAM), read-only memory (ROM), electrically erasable programmable read-only memory (EEPROM), programmable read-only memory (PROM), a magnetic memory, a magnetic disk, or an optical disk.

[0150] FIG. 10 is a block diagram of the device 100 according to an exemplary embodiment.

[0151] Referring to FIG. 10, the device 100 may include a sensor set 1010, a UI 1020, a memory 1030, a communicator 1040, an image processor 1050, an audio output transmitter 1060, an audio input receiver 1070, a camera 1080, and a processor 1090.

[0152] The device 100 may include a battery. The battery may be internally included in the device 100, or may be detachably included in the device 100. The battery may supply power to all elements included in the device 100. The device 100 may be supplied with power from an external power supply (not illustrated) via the communicator 1040. The device 100 may further include a connector that is connectable to the external power supply.

[0153] The processor 1090, the display 1021 and the user input receiver 1022 included in the UI 1020, the memory 1030, the audio input receiver 1070, and the communicator 1040, which are illustrated in FIG. 10, may be substantially similar or identical to the processor 930, the audio input receiver 910, the communicator 920, the display 940, the user input receiver 950, and the memory 960, which are illustrated in FIG. 9.

[0154] The programs stored in the memory 1030 may include a plurality of modules according to functions thereof. For example, the programs stored in the memory 1030 may include a UI module 1031, a notification module 1032, and an application module 1033, but are not limited thereto. For example, as in the memory 960 of FIG. 9, the programs stored in the memory 1030 may include a plurality of modules.

[0155] The UI module 1031 may provide the processor 1090 with a graphical UI (GUI) information for registering the wake-up keyword for speech recognition, GUI information indicating the speech recognition result (e.g., text information), and GUI information indicating a speech recognition waveform. The processor 1090 may display a screen on the display 1021 based on the GUI information received from the UI module 1031. The UI module 1031 may provide the processor 1090 with a UI and/or a GUI specialized with respect to each application installed on the device 100.

[0156] The notification module 1032 may provide a notification based on speech recognition, a notification

based on registration of the wake-up keyword, a notification based on incorrect input of the wake-up keyword, or a notification based on recognition of the wake-up keyword, but is not limited thereto.

[0157] The notification module 1032 may output a notification signal through the display 1021 as a video signal, or may output a notification signal through the audio output transmitter 1060 as an audio signal, but is not limited thereto.

[0158] The application module 1033 may include various applications, in addition to the speech recognition application described above.

[0159] The communicator 1040 may include one or more elements for communications between the device 100 and at least one external device (e.g., the speech recognition server 110, a smart TV, a smart watch, a smart mirror, and/or an IoT network based device, etc.). For example, the communicator 1040 may include at least one among a short-range wireless communicator 1041, a mobile communicator 1042, and a broadcasting receiver 1043, but is not limited thereto.

[0160] The short-range wireless communicator 1041 may include at least one among a Bluetooth communication module, a Bluetooth Low Energy (BLE) communication module, a near field communication (NFC) module, a wireless local area network (WLAN) (WiFi) communication module, a Zigbee communication module, an Ant+ communication module, a Wi-Fi Direction (WFD) communication module, a beacon communication module, and an Ultra WideBand (UWB) communication module, but is not limited thereto. For example, the short-range wireless communicator 1041 may include an infrared data association (IrDA) communication module.

[0161] The mobile communicator 1042 may transmit and receive a wireless signal with at least one among a base station, an external device, and a server via a mobile communication network. The wireless signal may include a voice call signal, a video call signal, or various types of data according to transmission and reception of text/multimedia messages.

[0162] The broadcasting receiver 1043 may receive broadcasting signals and/or broadcasting-related information from the outside via a broadcasting channel.

[0163] The broadcasting channel may include at least one among a satellite channel, a terrestrial channel, and a radio channel, but is not limited thereto.

[0164] In exemplary embodiments, the communicator 1040 may transmit at least one piece of information generated by the device 100 to at least one external device, or may receive information from at least one external device.

[0165] The sensor set 1010 may include a proximity sensor 1011 configured to sense an approach of the user 101 to the device 100, a bio sensor 1012, e.g., a heartbeat sensor, a blood flow meter, a diabetes sensor, a blood pressure sensor, and/or a stress sensor), configured to sense health information of the user 101 of the device 100, an illumination sensor 1013, e.g., a light-emitting

diode (LED) sensor, configured to sense ambient illumination of the device 100, a moodscope sensor 1014 configured to sense a mood of the user 101 of the device 100, an activity sensor 1015 configured to sense activity, a position sensor 1016, e.g., a GPS receiver, configured to detect a position of the device 100, a gyroscope sensor 1017 configured to measure an azimuth angle of the device 100, an accelerometer sensor 1018 configured to measure a slope and acceleration of the device 100 with respect to the surface of the earth, and/or a terrestrial magnetism sensor 1019 configured to sense azimuth orientation with respect to the device 100, but is not limited thereto.

[0166] For example, the sensor set 1010 may include a temperature/humidity sensor, a gravity sensor, an altitude sensor, a chemical sensor (e.g., an odorant sensor), a barometric pressure sensor, a fine dust measurement sensor, an ultraviolet sensor, an ozone sensor, a carbon dioxide (CO₂) sensor, and/or a network sensor (e.g., a network sensor based on Wi-Fi, Bluetooth, 3G, Long Term Evolution (LTE), and/or NFC), but is not limited thereto.

[0167] The sensor set 1010 may include a pressure sensor (e.g., a touch sensor, a piezoelectric sensor, a physical button, etc.), a state sensor (e.g., an earphone terminal, a DMB antenna, etc.), a standard terminal (e.g., a terminal capable of recognizing whether charging is in progress, a terminal capable of recognizing whether a PC is connected, a terminal capable of recognizing whether a dock is connected, etc.), and/or a time sensor, but is not limited thereto.

[0168] The sensor set 1010 may include less sensors than those illustrated in FIG. 10. For example, the sensor set 1010 may include only the position sensor 1016. In a case where the sensor set 1010 includes only the position sensor 1016, the sensor set 1010 may be referred to as a GPS receiver.

[0169] The results (or sensing values) sensed by the sensor set 1010 may be transmitted to the processor 1090. When the sensing value received from the sensor set 1010 is a value indicating a position, the processor 1090 may determine whether a current position of the device 100 is home or office, based on the received sensing value.

[0170] The processor 1090 may operate as a controller configured to an overall operation of the device 100. For example, the processor 1090 may control the sensor set 1010, the memory 1030, the UI 1020, the image processor 1050, the audio output transmitter 1060, the audio input receiver 1070, the camera 1080, and/or the communicator 1040 by executing the programs stored in the memory 1030.

[0171] The processor 1090 may operate in the same manner as the processor 930 of FIG. 9. For an operation of reading data from the memory 1030, the processor 1090 may perform an operation of receiving data from an external device via the communicator 1040. For an operation of writing data to the memory 1030, the proc-

essor 1090 may perform an operation of transmitting data to an external device via the communicator 1040.

[0172] The processor 1090 may perform at least one operation described above with reference to FIGS. 2, 3, and 4 to 8. The processor 1090 may be a controller configured to control the above-described operations.

[0173] The image processor 1050 may be configured to display image data received from the communicator 1040 or image data stored in the memory 1030 on the display 1021.

[0174] The audio output transmitter 1060 may output audio data received from the communicator 1040 or audio data stored in the memory 1030. The audio output transmitter 1060 may output an audio signal (e.g., a notification sound) related to the functions performed by the device 100.

[0175] The audio output transmitter 1060 may include a speaker or a buzzer, but is not limited thereto.

[0176] FIG. 11 is a functional block diagram of the speech recognition server 110 according to an exemplary embodiment.

[0177] Referring to FIG. 11, the speech recognition server 110 may include a communicator 1110, a processor 1120, and a memory 1130, but is not limited thereto.

The speech recognition server 110 may include less or more elements than those illustrated in FIG. 11.

[0178] The communicator 1110 may be substantially the same as the communicator 1040 illustrated in FIG. 10. The communicator 1110 may transmit a speech recognition related signal to the device 100 and receive a speech recognition related signal from the device 100.

[0179] The processor 1120 may perform the operation of the speech recognition server 110 described above with reference to FIGS. 2, 6, and 7.

[0180] The memory 1130 may store the wake-up keyword model 1131 and the speech recognition model 1132 and may provide the processor 1120 with the wake-up keyword model 1131 and the speech recognition model 1132 under the control of the processor 1120. The speech recognition model 1132 may be referred to as a model for recognizing the voice command.

[0181] The wake-up keyword model 1131 and the speech recognition model 1132, which are stored in the memory 1130, may be updated according to the information received via the communicator 1110. The wake-up keyword model 1131 and the speech recognition model 1132, which are stored in the memory 1130, may be updated according to information input by an operator. To this end, the speech recognition server 110 may further include an element configured to allow the operator to input information.

[0182] FIG. 12 is a configuration diagram of a speech recognition system 1200 according to an exemplary embodiment. FIG. 12 illustrates a case where the speech recognition server 110 recognizes speech signals of the user 101 that are received from a plurality of devices 1208.

[0183] The plurality of devices 1208 may include a mo-

mobile terminal 100, wearable glasses 1210, a smart watch 1220, an IoT device 1230, an IoT sensor 1240, and/or a smart TV 1250.

[0184] Users of the plurality of devices 1208 may be the same person or different persons. When the users of the plurality of devices 1208 are the same person, the speech recognition server 110 may register wake-up keyword models for each device, and perform the speech recognition function. When the users of the plurality of devices 1208 are different persons, the speech recognition server 110 may register wake-up keyword models by using device identification information and user identification information of each device, and perform the speech recognition function. Accordingly, the speech recognition system 1200 may provide various and more accurate speech recognition services. The speech recognition server 110 may provide the plurality of devices 1208 with the registered wake-up keyword models.

[0185] In addition, the speech recognition server 110 may estimate noise levels around the plurality of devices 1208 or recognize environment information by using the speech signal as well as the wake-up keyword according to the continuous recognition processing of the wake-up keyword and the voice command. The speech recognition server 110 may provide users with information used, estimated or recognized for controlling the plurality of devices 1208 by providing the plurality of devices 1208 with the estimated noise levels or the recognized environment information together with the speech recognition result.

[0186] The network 1260 may be a wired network and/or a wireless network. The network 1260 may allow data to be transmitted and received between the plurality of devices 1208 and the server 110 based on at least one among the communication methods described above in connection with the communicator 1040 illustrated in FIG. 10.

[0187] The methods described above with reference to FIGS. 2, 3, and 4 to 8 may be implemented by a computer program. For example, the operation of the device 100 illustrated in FIG. 2 may be performed by a speech recognition application installed on the device 100. The operation of the speech recognition server 110 illustrated in FIG. 2 may be performed by a speech recognition application installed on the speech recognition server 110. The computer program may operate in an OS environment installed on the device 100. The computer program may operate in an OS environment installed on the speech recognition server 110. The device 100 may write the computer program to a storage medium and may read the computer program from the storage medium. The speech recognition server 110 may write the computer program to a storage medium and may read the computer program from the storage medium.

[0188] According to an exemplary embodiment, the device 100 may comprise an audio input receiver 910 configured to receive an audio signal from a user, the audio signal 910 including a wake-up keyword, a memory 960 configured to store wake-up keyword models for

identifying the wake-up keyword from the received audio signal, and a processor 930 configured to perform detecting the wake-up keyword from the received audio signal by matching the wake-up keyword included into the received audio signal with the stored wake-up keyword models, generating a detection value indicative whether the wake-up keyword has been detected or not, based on a result of the matching, transmitting the detection value and the received audio signal to a server, receiving, from the server, a speech recognition result of the audio signal that is translated based on the detection value, and controlling an executable application of the device in performing a device function, based on the speech recognition result.

[0189] The detection value indicates that the wake-up keyword has been detected in the received audio signal, and the processor 930 is configured to receive the speech recognition result which includes a user command to execute the application, and in which the wake-up keyword itself is absent.

[0190] The audio input receiver 910 is configured to receive, in advance, individual user inputs containing individual keywords related to a control of executable applications of the device 100, and the memory 960 is configured to store the wake-up keyword models based on the received individual keywords.

[0191] According to an exemplary embodiment, a method may comprise storing, in a first memory, wake-up keyword models which are used to identify wake-up keywords; receiving an audio signal from a user, the audio signal including the wake-up keyword; detecting the wake-up keyword from the received audio signal by matching the wake-up keyword included into the received audio signal with the stored wake-up keyword models, generating a detection value indicative whether the wake-up keyword has been detected or not based on a result of the matching; transmitting the detection value and the received audio signal to a server, receiving, from the server, a speech recognition result of the audio signal that is translated based on the detection value, and controlling an executable application of the device in performing a device function, based on the speech recognition result.

[0192] The method further comprises storing, in a second memory, a speech recognition model which is used to translate the audio signal of the user and the wake-up keyword models which are synchronized with the wake-up keyword models stored in the first memory, wherein the receiving the speech recognition result comprises: identifying, by the server, whether the audio signal contains the wake-up keyword, from the detection value; and translating, by the server, the audio signal into the speech recognition result based on a combination model in which the speech recognition model is combined with a respective wake-up keyword model in response to the detection value indicating that the audio signal contains the wake-up keyword. The first memory and the second memory may be included in the memory 960.

[0193] The receiving the speech recognition result further comprises generating, by the server, the speech recognition result by removing the wake-up keyword from the speech recognition result, and receiving, from the server 110, the speech recognition result of the audio signal from which the wake-up keyword has been removed; wherein the controlling comprises controlling the executable application of the device 100 according to the speech recognition result from which the wake-up keyword has been removed.

[0194] The translating comprises translating the audio signal into the speech recognition result by using the speech recognition model only, in response to the detection value indicating that the audio signal does not contain the wake-up keyword.

[0195] Exemplary embodiments may be embodied in a storage medium including instruction codes executable by a computer, such as a program module executed by the computer. A computer-readable medium may be any available medium which can be accessed by the computer and may include any volatile/non-volatile media and any removable/non-removable media. Furthermore, the computer-readable medium may include any computer storage and communication media. The computer storage medium may include any volatile/non-volatile and removable/non-removable media embodied by a certain method or technology for storing information such as computer-readable instruction code, a data structure, a program module or other data. The communication media may include computer-readable instruction code, a data structure, a program module, other data of modulated data signals, or other transmission mechanisms, and may include any information transmission media.

[0196] The foregoing exemplary embodiments and advantages are merely exemplary and are not to be construed as limiting. The present teaching can be readily applied to other types of apparatuses. Also, the description of the exemplary embodiments is intended to be illustrative, and not to limit the scope of the claims, and many alternatives, modifications, and variations will be apparent to those skilled in the art.

Claims

1. A system (10) comprising: a device (100) and a speech recognition server (110), the device comprising:

- an audio input receiver (910) configured to receive a speech signal of a user;
- a memory (960) configured to store a wake-up keyword model;
- a communicator (920) configured to communicate with the speech recognition server;
- a processor (930) configured to :
 - detect a wake-up keyword from the speech signal of the user by using the wake-up keyword

model in response to receiving the speech signal of the user (S605),

characterized in that

the processor is further configured to:

transmit a wake-up keyword detection signal and the speech signal of the user to the speech recognition server via the communicator if the wake-up keyword is detected in the speech signal by the processor (S606),
 transmit a wake-up keyword non-detection signal and the speech signal of the user to the speech recognition server via the communicator if the wake-up keyword is not detected in the speech signal by the processor (S606), receive, from the speech recognition server via the communicator, a speech recognition result of the speech signal based on the wake-up keyword detection/non-detection signal (S610),
 the speech recognition result being generated by the speech recognition server :

using a speech recognition model,
 wherein the speech recognition model is combined with the wake-up keyword model if the wake-up keyword is detected,
 and wherein the speech recognition model is not combined with the wake-up keyword model if the wake-up keyword is not detected,
 and control the device according to the speech recognition result (611);

the device further comprising :

an environment information sensor configured to sense environment information about the device,
 wherein the wake-up keyword model is one of a plurality of wake-up keyword models which are associated with various environment information, and the processor is further configured to detect the wake-up keyword from the received speech signal of the user by using the wake-up keyword model that is associated with the sensed environment information of the device, among the plurality of wake-up keyword models (S605); and

the device is configured to :

register the wake-up keyword models based on the environment information; send the environmental information to the server with the speech signal and wake-up detection/non detection signal;
 so that the server can then :

- generate the speech recognition result using a method comprising the step of removing the keyword from the received speech signal in dependence of a register of wake-up keyword models, stored on the server, and based on the environment information; and transmit the speech recognition result to device.
- 2. The system of claim 1, wherein the environment information includes position information of the device.
- 3. The system of claim 1, wherein the processor of the device is further configured to acquire identification information of the user based on the received speech signal of the user (S705), transmit the identification information of the user and the environment information of the device to the speech recognition server via the communicator (S707), and receive, from the speech recognition server, the speech recognition result based on at least two of the wake-up keyword detection/non-detection signal, identification information of the user, and the environment information of the device (S711).
- 4. The system of claim 1, wherein when an audio signal for registering the wake-up keyword model is received through the audio input receiver, the processor is further configured to determine whether the audio signal is valid as the wake-up keyword model, based on a matching rate of a recognition of the audio signal.
- 5. The system of claim 1, the device further comprising a user input receiver configured to receive a user input, wherein the memory is configured to store candidate wake-up keyword models, and the processor is further configured to register a candidate wake-up keyword model, which is selected among the stored candidate wake-up keyword models, according to the user input received through the user input receiver.
- 6. The system of claim 1, wherein the processor of the device is further configured to receive the wake-up keyword model from the speech recognition server via the communicator and register the received wake-up keyword model in the memory.
- 7. The system of claim 1, wherein the processor of the device is further configured to generate the wake-up keyword model based on the speech signal of the user, which is received through the audio input receiver, and register the generated wake-up keyword model in the memory.

- 8. A speech recognition server (110) comprising:
 - a communicator (1110) configured to communicate with a device;
 - a memory (1130) configured to store a wake-up keyword model and a speech recognition model; and
 - a processor (1120),
 characterized in that:
 - the processor is configured to receive a speech signal of a user and a wake-up keyword detection signal from the device via the communicator if the wake-up keyword is detected in the speech signal by the device (S606), receive the speech signal of the user and a wake-up keyword non-detection signal from the device via the communicator if the wake-up keyword is not detected in the speech signal by the device (S606), set a speech recognition model combined with the wake-up keyword model in response to receiving the wake-up keyword detection signal (S607), set a speech recognition model that is not combined with the wake-up keyword model in response to receiving the wake-up keyword non-detection signal (S607), recognize the speech signal of the user by using the set speech recognition model (S608) and generate a speech recognition result of the speech signal, remove the wake-up keyword from the speech recognition result of the speech signal of the user (S609), and transmit the speech recognition result, from which the wake-up keyword has been removed, to the device via the communicator (S610);
 - wherein the wake-up keyword model is one of a plurality of wake-up keyword models which are associated with various environment information, and
 - the processor is further configured to set the speech recognition model combined with the wake-up keyword model that is associated with the environment information of the device, among the plurality of wake-up keyword models, in response to receiving the environment information of the device via the communicator (S607), and generate the speech recognition result using a method comprising the step of removing the keyword from the received speech signal in dependence of a register of wake-up keyword models, stored on the server, and based on the environment information (S609).
- 9. The speech recognition server of claim 8, wherein the processor is further configured to set the speech recognition model combined with identification information of the user and the wake-up keyword model

corresponding to the environment information of the device (S708), in response to receiving the identification information of the user via the communicator, and generate the speech recognition result using the speech recognition model, wherein the identification information includes an identifying characteristic of the user.

10. A speech recognition method, which is performed by a device and a server, the speech recognition method comprising:

receiving a speech signal of a user (S603);
 detecting a wake-up keyword from the received speech signal of the user by using a wake-up keyword model (S605);
characterized by the method further comprising:

transmitting a wake-up keyword detection signal and the speech signal of the user to a speech recognition server if the wake-up keyword is detected in the speech signal (S606),

transmitting a wake-up keyword non-detection signal and the speech signal of the user to the speech recognition server if the wake-up keyword is not detected in the speech signal (S606);

receiving a result of recognizing the speech signal of the user from the speech recognition server, based on the wake-up keyword detection/non-detection signal and sensed environment information sensed by an environment information sensor of the device (S610); and

controlling the device according to the result of recognizing the speech signal of the user (S611),

wherein the wake-up keyword model is one of a plurality of wake-up keyword models which are associated with various environment information, and wherein detecting the wake-up keyword from the received speech signal of the user comprises using the wake-up keyword model that is associated with the sensed environment information of the device, among the plurality of wake-up keyword models (S605); and
 wherein the method further comprises registering the wake-up keyword models based on the environment information (S601);
 sending the environmental information to the server with the speech signal and wake-up detection/non detection signal (S606);
 so that the server can then :

generate the speech recognition result us-

ing a method comprising the step of removing the keyword from the received speech signal in dependence of a register of wake-up keyword models, stored on the server, and based on the environment information; and
 transmit the speech recognition result to the device.

11. A speech recognition method, which is performed by a speech recognition server, **characterized by** the speech recognition method comprising:

receiving, from a device, a wake-up keyword detection signal and a speech signal of a user if a wake-up keyword is detected in the speech signal by the device (S606);

receiving, from the device, a wake-up keyword non-detection signal and the speech signal of the user if the wake-up keyword is not detected in the speech signal by the device (S606);

setting a speech recognition model combined with a wake-up keyword model in response to receiving the wake-up keyword detection signal (S607);

setting a speech recognition model that is not combined with the wake-up keyword model in response to receiving the wake-up keyword non-detection signal (S607);

recognizing the speech signal of the user by using the set speech recognition model (S608) and generating a speech recognition result of the speech signal;

removing the wake-up keyword from the speech recognition result (S609); and

transmitting the speech recognition result, from which the wake-up keyword has been removed, to the device (S610),

wherein the wake-up keyword model is one of a plurality of wake-up keyword models which are associated with various environment information, and

wherein setting the speech recognition model combined with the wake-up keyword model comprises setting the speech recognition model combined with the wake-up keyword model that is associated with the environment information of the device, among the plurality of wake-up keyword models, in response to receiving the environment information of the device (S607), and

wherein generating the speech recognition result comprises using a method comprising the step of removing the keyword from the received speech signal in dependence of a register of wake-up keyword models, stored on the server, and based on the environment information (S609).

12. A non-transitory computer-readable recording medium having stored thereon a program which, when executed by a computer system, causes the computer system to perform the method of claim 10.

5

13. A non-transitory computer-readable recording medium having stored thereon a program which, when executed by a computer system, causes the computer system to perform the method of claim 11.

10

Patentansprüche

1. Ein System (10), das Folgendes umfasst: Ein Gerät (100) und einen Spracherkennungsserver (110), wobei das Gerät Folgendes umfasst:

15

einen Audioeingangsempfänger (910), der ein Sprachsignal eines Nutzers empfängt, einen Speicher (960) der ein Weck-Schlüsselwortmuster speichert;

20

einen Kommunikator (920), der mit dem Spracherkennungsserver kommuniziert;

einen Prozessor (930), der:

ein Weck-Schlüsselwort von dem Sprachsignal des Nutzers erkennt, indem er das Weck-Schlüsselwortmuster als Reaktion auf den Empfang des Sprachsignals des Nutzers (S605) verwendet,

25

dadurch charakterisiert, dass der Prozessor weiterhin so konfiguriert ist, dass er:

30

ein Erkennungssignal von dem Weck-Schlüsselwort und das Sprachsignal des Nutzers an den Spracherkennungsserver über den Kommunikator übermittelt, wenn das Weck-Schlüsselwort von dem Prozessor (S606) in dem Sprachsignal erkannt wird, ein Nicht-Erkennungssignal von dem Weck-Schlüsselwort und das Sprachsignal des Nutzers an den Spracherkennungsserver über den Kommunikator übermittelt, wenn das Weck-Schlüsselwort nicht von dem Prozessor (S606) in dem Sprachsignal erkannt wird;

35

40

45

und von dem Spracherkennungsserver über den Kommunikator ein Spracherkennungsergebnis, basierend auf dem Erkennungs-/Nicht-Erkennungssignal (S610) des Weck-Schlüsselworts empfängt,

50

wobei das Spracherkennungsergebnis von dem Spracherkennungsserver generiert wird:

55

durch Verwendung eines Erkennungsmusters, wobei das Spracherkennungsmuster mit

dem Weck-Schlüsselwortmuster kombiniert wird, wenn das Weck-Schlüsselwort erkannt wird, und wobei das Spracherkennungsmuster nicht mit dem Weck-Schlüsselwortmuster kombiniert wird, wenn das Weck-Schlüsselwort nicht erkannt wird, und durch Steuern des Geräts gemäß dem Spracherkennungsergebnisses (611);

wobei das Gerät des Weiteren Folgendes umfasst:

einen Umgebungsinformationssensor, der Umgebungsinformation über das Gerät spürt, wobei das Weck-Schlüsselwortmuster eines von mehreren Weck-Schlüsselwortmustern ist, die mit verschiedenen Umgebungsinformationen assoziiert sind, und der Prozessor weiterhin so konfiguriert ist, dass er das Weck-Schlüsselwort von dem empfangenen Sprachsignal des Nutzers erkennt, indem er unter mehreren Weck-Schlüsselwortmustern das Weck-Schlüsselwortmuster verwendet, das mit der gespürten Umgebungsinformation assoziiert ist (S605); und das Gerät weiterhin so konfiguriert ist, dass es:

die Weck-Schlüsselwortmuster basierend auf der Umgebungsinformation registriert; die Umgebungsinformation mit dem Sprachsignal und dem Erkennungs-/Nicht-Erkennungssignal an den Server sendet;

sodass der Server dann das Spracherkennungsergebnis generieren kann, durch Anwenden einer Methode, die darin besteht, das Schlüsselwort des empfangenen Sprachsignals abhängig von einem auf dem Server gespeicherten Register der Weck-Schlüsselwortmuster und basierend auf der Umgebungsinformation zu entfernen; und das Spracherkennungsergebnis an das Gerät übermittelt.

2. Das System aus Anspruch 1, wobei die Umgebungsinformation Information über die Position des Geräts umfasst.

3. Das System aus Anspruch 1, wobei der Prozessor des Geräts des Weiteren so konfiguriert ist, dass er Identifizierungsinformation des Nutzers basierend auf dem empfangenen Sprachsignal des Nutzers (S705) erhält, die Identifizierungsinformation des Nutzers und die Umgebungsinformation des Geräts an den Spracherkennungsserver über den Kommu-

- nikator (S707) übermittelt und von dem Spracherkennungsserver das Spracherkennungsergebnis basierend auf mindestens zwei der Erkennungs-/Nicht-Erkennungssignale des Weck-Schlüsselworts, der Identifizierungsinformation des Nutzers und der Umgebungsinformation des Geräts (S711) empfängt.
4. Das System aus Anspruch 1, wobei, wenn ein Audiosignal für das Registrieren des Weck-Schlüsselwortmusters über den Audioeingangsempfänger empfangen wird und der Prozessor des Weiteren so konfiguriert ist, dass er bestimmt, ob das Audiosignal als Weck-Schlüsselwortmuster gültig ist, basierend auf einer Übereinstimmungsquote des Audiosignals.
5. Das System aus Anspruch 1, wobei das Gerät des Weiteren einen Nutzereingangsempfänger umfasst, der einen Nutzereingang empfängt, wobei der Speicher so konfiguriert ist, dass er mögliche Weck-Schlüsselwortmuster speichert und der Prozessor weiterhin so konfiguriert ist, dass er ein mögliches Weck-Schlüsselwortmuster registriert, welches er aus den gespeicherten möglichen Weck-Schlüsselwortmustern wählt, gemäß dem über den Nutzereingangsempfänger empfangenen Nutzereingang.
6. Das System aus Anspruch 1, wobei der Prozessor des Geräts des Weiteren so konfiguriert ist, dass er das Weck-Schlüsselwortmuster von dem Spracherkennungsserver über den Kommunikator empfängt und das empfangene Weck-Schlüsselwortmuster in dem Speicher registriert.
7. Das System aus Anspruch 1, wobei der Prozessor des Geräts des Weiteren so konfiguriert ist, dass er das Weck-Schlüsselwortmuster basierend auf dem Sprachsignal des Nutzers, das über den Audioeingangsempfänger empfangen wird, generiert und das generierte Weck-Schlüsselwortmuster in dem Speicher registriert.
8. Ein Spracherkennungsserver (110), der Folgendes umfasst:
- einen Kommunikator (1110), der mit einem Gerät kommuniziert;
- einen Speicher (1130), der ein Weck-Schlüsselwortmuster und ein Spracherkennungsmuster speichert; und
- einen Prozessor (1120), der sich dadurch auszeichnet, dass:
- er ein Sprachsignal von einem Nutzer und ein Erkennungssignal von einem Weck-Schlüsselwort von dem Gerät über den

5

10

15

20

25

30

35

40

45

50

55

Kommunikator empfängt, wenn das Weck-Schlüsselwort von dem Gerät (S606) in dem Sprachsignal erkannt wird, und das Sprachsignal des Nutzers und ein Nicht-Erkennungssignal von einem Weck-Schlüsselwort von dem Gerät über den Kommunikator empfängt, wenn das Weck-Schlüsselwort in dem Sprachsignal von dem Gerät (S606) nicht erkannt wird, ein Spracherkennungsmuster, kombiniert mit dem Weck-Schlüsselwortmuster als Reaktion auf das Erkennungssignal des Weck-Schlüsselworts (S607) festlegt, ein Spracherkennungsmuster, nicht kombiniert mit dem Weck-Schlüsselwortmuster als Reaktion auf den Empfang des Nicht-Erkennungssignals des Weck-Schlüsselworts (S607) festlegt, das Sprachsignal des Nutzers durch Verwendung des festgelegten Spracherkennungsmusters (S608) erkennt und ein Spracherkennungsergebnis des Sprachsignals generiert, das Weck-Schlüsselwort aus dem Spracherkennungsergebnis des Sprachsignals des Nutzers (S609) entfernt und das Spracherkennungsergebnis, von dem das Weck-Schlüsselwort entfernt wurde, über den Kommunikator an das Gerät (S610) sendet;

wobei das Weck-Schlüsselwortmuster eines von mehreren Weck-Schlüsselwortmustern ist, die mit verschiedenen Umgebungsinformationen assoziiert sind, und der Prozessor weiterhin so konfiguriert ist, dass er das

Spracherkennungsmuster festlegt, kombiniert mit dem Weck-Schlüsselwortmuster, das mit der Umgebungsinformation des Geräts assoziiert ist, wobei er aus mehreren Weck-Schlüsselwortmustern auswählt, als Reaktion auf den Empfang der Umgebungsinformation des Geräts über den Kommunikator (S607), und das Spracherkennungsergebnis generiert, indem er eine Methode verwendet, die das Entfernen des Schlüsselworts von dem empfangenen Sprachsignal, abhängig von einem auf dem Server gespeicherten Register von Weck-Schlüsselwortmustern umfasst, und basierend auf der Umgebungsinformation (S609).

9. Der Spracherkennungsserver aus Anspruch 8, wobei der Prozessor weiterhin so konfiguriert ist, dass er das Spracherkennungsmuster festlegt, kombiniert mit Identifizierungsinformation des Nutzers und dem Weck-Schlüsselwortmuster gemäß der Umgebungsinformation des Geräts (S708), als Reaktion auf den Empfang der Identifizierungsinformation des Nutzers über den Kommunikator, und das Spracher-

kennungsergebnis generiert, indem er das Spracherkennungsmuster verwendet, wobei die Identifizierungsinformation ein bestimmtes Merkmal des Nutzers umfasst.

10. Eine Spracherkennungsmethode, die von einem Gerät und einem Server ausgeführt wird, wobei die Spracherkennungsmethode Folgendes umfasst:

den Empfang eines Sprachsignals eines Nutzers (S603);
 das Erkennen eines Weck-Schlüsselworts von dem empfangenen Sprachsignal des Nutzers durch Verwenden eines Weck-Schlüsselwortmusters (S605);
 wobei die Methode des Weiteren Folgendes umfasst:

die Übermittlung eines Erkennungssignals von dem Weck-Schlüsselwort und des Sprachsignals des Nutzers an den Spracherkennungsserver, wenn das Weck-Schlüsselwort von dem Prozessor in dem Sprachsignal (S606) erkannt wird,
 die Übermittlung eines Nicht-Erkennungssignals von einem Weck-Schlüsselwort und des Sprachsignals des Nutzers an den Spracherkennungsserver, wenn das Weck-Schlüsselwort nicht von dem Prozessor in dem Sprachsignal (S606) erkannt wird,
 der Empfang seitens des Spracherkennungsservers eines Ergebnisses einer Nutzer-Sprachsignalerkennung, basierend auf dem Erkennungs-/Nicht-Erkennungssignal von dem Weck-Schlüsselwort und von Umgebungsinformation, die ein Umgebungsinformationssensor des Geräts (S610) spürt; und
 das Steuern des Geräts gemäß dem Ergebnis der Nutzer-Sprachsignalerkennung (S611),

wobei das Weck-Schlüsselwortmuster eines von mehreren Weck-Schlüsselwortmustern ist, die mit den verschiedenen Umgebungsinformationen assoziiert sind, und wobei das Erkennen des Weck-Schlüsselworts von dem empfangenen Sprachsignal des Nutzers umfasst, dass er aus mehreren Weck-Schlüsselwortmustern das Weck-Schlüsselwortmuster auswählt, das mit der gespürten Umgebungsinformation des Geräts assoziiert ist (S605); und
 wobei die Methode des Weiteren das Registrieren des Weck-Schlüsselwortmusters basierend auf der Umgebungsinformation (S601) umfasst; die Umgebungsinformation an den Server sendet, zusammen mit dem Sprachsignal und dem Erkennungs-/Nicht-Erkennungssignal (S606)

von dem Weck-Schlüsselwort; sodass der Server dann

das Spracherkennungsergebnis generieren kann, durch Anwenden einer Methode, die darin besteht, das Schlüsselwort des empfangenen Sprachsignals abhängig von einem auf dem Server gespeicherten Register der Weck-Schlüsselwortmuster und basierend auf der Umgebungsinformation zu entfernen; und das Spracherkennungsergebnis an das Gerät übermittelt.

11. Eine Spracherkennungsmethode, die von einem Spracherkennungsserver ausgeführt wird und sich dadurch auszeichnet, dass sie des Weiteren Folgendes umfasst:

den Empfang, von einem Gerät, eines Erkennungssignals von einem Weck-Schlüsselwort und einem Sprachsignal eines Nutzers, wenn das Gerät (S606) ein Weck-Schlüsselwort in dem Sprachsignal erkennt;
 den Empfang, von dem Gerät, eines Nicht-Erkennungssignals von dem Weck-Schlüsselwort und des Sprachsignals des Nutzers, wenn das Gerät (S606) das Weck-Schlüsselwort nicht in dem Sprachsignal erkennt;
 das Festlegen eines Spracherkennungsmusters, kombiniert mit einem Weck-Schlüsselwortmuster, als Reaktion auf den Empfang des Erkennungssignals des Weck-Schlüsselworts (S607);
 das Festlegen eines Spracherkennungsmusters, nicht kombiniert mit einem Weck-Schlüsselwortmuster, als Reaktion auf den Empfang des Nicht-Erkennungssignals des Weck-Schlüsselworts (S607);
 das Erkennen des Sprachsignals des Nutzers durch Verwenden des festgelegten Spracherkennungsmusters (S608) und das Generieren eines Spracherkennungsergebnisses des Sprachsignals;
 das Entfernen des Weck-Schlüsselworts von dem Spracherkennungsergebnis (S609); und
 das Übermitteln des Spracherkennungsergebnisses, von dem das Weck-Schlüsselwort entfernt wurde, an das Gerät (S610),
 wobei das Weck-Schlüsselwortmuster eines von mehreren Weck-Schlüsselwortmustern ist, die mit verschiedenen Umgebungsinformationen assoziiert sind, und
 wobei das Festlegen des Spracherkennungsmusters, kombiniert mit dem Weck-Schlüsselwortmuster, das Festlegen des Spracherkennungsmusters kombiniert mit dem Weck-Schlüsselwortmuster, das mit der Umgebungsinformation des Geräts assoziiert ist, umfasst, als Reaktion auf den Empfang der Umgebungs-

information des Geräts (S607), und wobei das Generieren des Spracherkennungsergebnisses das Anwenden einer Methode umfasst, die darin besteht, das Schlüsselwort des empfangenen Sprachsignals abhängig von einem auf dem Server gespeicherten Register der Weck-Schlüsselwortmuster und basierend auf der Umgebungsinformation (S609) zu entfernen.

12. Ein nichtflüchtiges computerlesbares Speichermedium, auf dem ein Programm gespeichert ist, das, wenn es von einem Computer ausgeführt wird, diesen dazu bringt, die Methode aus Anspruch 10 auszuführen.

13. Ein nichtflüchtiges computerlesbares Speichermedium, auf dem ein Programm gespeichert ist, das, wenn es von einem Computer ausgeführt wird, diesen dazu bringt, die Methode aus Anspruch 11 auszuführen.

Revendications

1. Système (10) incluant un dispositif (100) et un serveur de reconnaissance vocale (110), le dispositif comprenant :

un récepteur audio (910) configuré pour recevoir un signal vocal d'un utilisateur ; une mémoire (960) configurée pour stocker un modèle d'éveil par mot-clé ; un communicateur (920) configuré pour communiquer avec le serveur de reconnaissance vocale ; un processeur (930) configuré pour :

détecter un mot-clé d'éveil dans le signal vocal de l'utilisateur en utilisant le modèle d'éveil par mot-clé en réponse à la réception du signal vocal de l'utilisateur (S605), et **caractérisé en ce que**

le processeur est aussi configuré pour :

transmettre le signal de détection par mot-clé d'éveil et le signal vocal de l'utilisateur au serveur de reconnaissance vocale via le communicateur lorsque le mot-clé d'éveil est détecté dans le signal vocal par le processeur (S606),

transmettre un signal de non détection par mot-clé d'éveil et le signal vocal de l'utilisateur au serveur de reconnaissance vocale via le communicateur si le mot-clé d'éveil n'est pas détecté dans le signal vocal par le processeur (S606), recevoir (du serveur de reconnaissance vo-

cale via le communicateur) le résultat de la reconnaissance vocale du signal vocal basé sur les signaux de détection et de non détection par mot-clé d'éveil (S610),

le résultat de la reconnaissance vocale étant généré par le serveur de reconnaissance vocale :

selon un modèle de reconnaissance vocale, où le modèle de reconnaissance vocale est combiné avec le modèle d'éveil par mot-clé lorsque le mot-clé d'éveil est détecté, et où le modèle de reconnaissance vocale n'est pas combiné avec le modèle d'éveil par mot-clé si le mot-clé d'éveil n'est pas détecté, et le dispositif est contrôlé en fonction du résultat de la reconnaissance vocale (611).

Le dispositif inclut en outre :

un détecteur d'informations environnementales configuré pour détecter des informations environnementales sur le dispositif, où le modèle d'éveil par mot-clé fait partie de plusieurs modèles d'éveil par mot-clé associés à différentes informations environnementales. Le processeur est aussi configuré pour détecter le mot-clé d'éveil contenu dans le signal vocal de l'utilisateur en utilisant le modèle d'éveil par mot-clé qui est associé aux informations environnementales détectées du dispositif, parmi plusieurs modèles d'éveil par mot-clé (S605) ; et

le dispositif est configuré pour :

enregistrer les modèles d'éveil par mot-clé basés sur des informations environnementales, envoyer les informations environnementales au serveur avec le signal vocal et le signal de détection/non-détection de détection d'éveil ; de telle façon que le serveur puisse :

générer le résultat de la reconnaissance vocale en utilisant une méthode qui combine l'effacement du mot-clé dans le signal vocal reçu en fonction d'une liste des modèles d'éveil par mot-clé, stockés dans le serveur, et basés sur des informations environnementales et ; transmettre le résultat de la reconnaissance vocale au dispositif.

2. Le système de la revendication 1, dont les informations environnementales incluent des informations

de position du dispositif.

3. Le système de la revendication 1, dont le processeur du dispositif est configuré pour acquérir les informations d'identification de l'utilisateur basées sur le signal vocal reçu de l'utilisateur (S705), transmettre les informations d'identification de l'utilisateur et les informations environnementales du dispositif au serveur de reconnaissance vocale via le communicateur (S707), et recevoir (du serveur de reconnaissance vocale) le résultat de la reconnaissance vocale basé sur au moins deux signaux de détection/non-détection mot-clé d'éveil, les informations d'identification de l'utilisateur, et les informations environnementales du dispositif (S711) .
4. Dans le système de la revendication 1, un signal audio d'enregistrement du modèle d'éveil par mot-clé est reçu par le récepteur audio. Le processeur est également configuré pour déterminer si le signal audio est valide comme modèle d'éveil par mot-clé, en fonction du taux de correspondance de la reconnaissance du signal audio.
5. Dans le système de la revendication 1, le dispositif inclut également un récepteur d'entrée d'utilisateur configuré pour recevoir une entrée d'utilisateur, dont la mémoire est configurée pour stocker des modèles d'éveil par mot-clé candidats et dont le processeur est aussi configuré pour enregistrer le modèle d'éveil par mot-clé candidat sélectionné parmi les modèles d'éveil par mot-clé candidats, en fonction des entrées d'utilisateur reçues par le récepteur des entrées d'utilisateur.
6. Dans le système de la revendication 1, le processeur du dispositif est configuré pour recevoir le modèle d'éveil par mot-clé du serveur de reconnaissance vocale via le communicateur et pour enregistrer dans la mémoire le modèle d'éveil par mot-clé reçu.
7. Dans le système de la revendication 1, le processeur du dispositif est configuré pour générer le modèle d'éveil par mot-clé basé sur le signal vocal de l'utilisateur, qui est reçu par le récepteur audio, et pour enregistrer dans la mémoire le modèle d'éveil par mot-clé généré.
8. Un serveur de reconnaissance vocale (110) qui inclut :
 - un communicateur (1110) configuré pour communiquer avec un dispositif ;
 - une mémoire (1130) configurée pour stocker un modèle d'éveil par mot-clé et un modèle de reconnaissance vocale, et
 - un processeur (1120) configuré pour :

recevoir un signal vocal d'un utilisateur et un signal de détection de mot-clé d'éveil provenant du dispositif via le communicateur lorsque le mot-clé d'éveil est détecté dans le signal vocal émis par le dispositif (S606),

recevoir le signal vocal de l'utilisateur et un signal de non-détection de mot-clé d'éveil du dispositif via le communicateur lorsque le mot clé d'éveil n'est pas détecté dans le signal vocal par le dispositif (S606),

définir un modèle de reconnaissance vocale combiné avec le modèle d'éveil par mot-clé en réponse à la réception du mot-clé d'éveil signal de détection (S607),

définir aussi un modèle de reconnaissance vocale qui n'est pas combiné avec le modèle d'éveil par mot-clé en réponse au signal de non-détection de mot-clé d'éveil (S607),

reconnaître le signal vocal de l'utilisateur en utilisant le modèle de reconnaissance vocale (S608) et générer le résultat de la reconnaissance vocale du signal vocal, avec effacement du mot clé d'éveil dans le résultat de la reconnaissance vocale du signal vocal de l'utilisateur (S609),

et transmettre le résultat de la reconnaissance vocale, dans lequel le mot-clé d'éveil a été effacé, jusqu'au dispositif via le communicateur (S610). Le modèle d'éveil par mot-clé fait partie des modèles d'éveil par mot-clé qui sont associés aux diverses informations environnementales, et

le processeur est aussi configuré pour définir le modèle de reconnaissance vocale combiné avec le modèle d'éveil par mot-clé qui est associé avec les informations environnementales du dispositif, parmi les différents modèles d'éveil par mot-clé, en réponse à la réception des informations environnementales du dispositif via le communicateur (S607), et pour générer le résultat de la reconnaissance vocale en utilisant une méthode qui inclut l'effacement du mot clé dans le signal vocal en fonction d'un registre de modèles d'éveil par mot-clé, stockés sur le serveur et basés sur les informations environnementales (S609).

9. Dans le serveur de reconnaissance vocale de la revendication 8, le processeur est aussi configuré pour définir le modèle de reconnaissance vocale combiné avec les informations d'identification de l'utilisateur et le modèle d'éveil par mot-clé correspondant aux informations environnementales du dispositif (S708), en réponse à la réception des informations d'identification de l'utilisateur via le communicateur,

et pour générer le résultat de la reconnaissance vocale en utilisant le modèle de reconnaissance vocale, dont les informations d'identification incluent des caractéristiques d'identification de l'utilisateur.

- 5
10. Une méthode de reconnaissance vocale, exécutée par un dispositif et un serveur, la méthode de reconnaissance vocale incluant :

la réception d'un signal vocal d'un utilisateur (S603) ;

la détection d'un mot-clé d'éveil dans le signal vocal reçu de l'utilisateur conformément à un modèle d'éveil par mot-clé (S605) ;

15 **caractérisé par** la méthode qui inclut également :

la transmission d'un signal de détection par mot-clé d'éveil et le signal vocal de l'utilisateur vers un serveur de reconnaissance vocale lorsque le mot-clé d'éveil est détecté dans le signal vocal (S606),

la transmission d'un signal de non-détection de mot-clé d'éveil et le signal vocal de l'utilisateur envoyé au serveur de reconnaissance vocale lorsque le mot-clé d'éveil n'est pas détecté dans le signal vocal (S606) ;

la réception d'un résultat de reconnaissance du signal vocal de l'utilisateur provenant du serveur de reconnaissance vocale, basé sur le signal de détection/ non-détection du mot-clé d'éveil et les informations environnementales détectées par un détecteur d'informations environnementales du dispositif (S610) et

le contrôle du dispositif suivant le résultat de la reconnaissance du signal vocal de l'utilisateur (S611),

où le modèle d'éveil par mot-clé fait partie des modèles d'éveil par mot-clé qui sont associés à différentes informations environnementales et où la détection du mot-clé d'éveil dans le signal vocal reçu de l'utilisateur est basée sur le modèle d'éveil par mot-clé qui est associé aux informations environnementales détectées du dispositif, parmi les différents modèles d'éveil par mot-clé (S605) ; et

où la méthode inclut en outre l'enregistrement des modèles d'éveil par mot-clé basés sur les informations environnementales (S601) ;

pour envoyer les informations environnementales au serveur avec le signal vocal et le signal de détection/ non-détection de mot-clé d'éveil (S606), permettant alors au serveur de générer le résultat de la reconnaissance vocale en utilisant une méthode qui inclut l'effacement du mot-clé dans le signal vocal reçu conformément au

registre des modèles de mots-clés d'éveil stockés sur le serveur, et basés sur les informations environnementales, et transmettre le résultat de la reconnaissance vocale au dispositif.

11. Une méthode de reconnaissance vocale, qui est exécutée par un serveur de reconnaissance vocale, avec la méthode de reconnaissance vocale incluant :

la réception d'un signal de détection de mot-clé d'éveil et un signal vocal d'un utilisateur lorsqu'un mot-clé d'éveil est détecté dans le signal vocal par le dispositif (S606) ;

la réception d'un signal (provenant du dispositif) de non-détection de mot-clé d'éveil et du signal vocal de l'utilisateur si le mot-clé d'éveil n'est pas détecté dans le signal vocal par le dispositif (S606) ;

la définition d'un modèle de reconnaissance vocale combiné à un modèle d'éveil par mot-clé en réponse à la réception d'un signal de détection de mot-clé d'éveil (S607) ;

la définition d'un modèle de reconnaissance vocale qui n'est pas combiné au modèle d'éveil par mot-clé en réponse à la réception d'un signal de non-détection de mot-clé d'éveil (S607) ;

la reconnaissance du signal vocal de l'utilisateur en utilisant le modèle de reconnaissance vocale défini (S608) et la génération d'un résultat de la reconnaissance vocale du signal vocal ;

l'effacement du mot-clé d'éveil dans le résultat de la reconnaissance vocale (S609) ; et la transmission au dispositif du résultat de la reconnaissance vocale, dans lequel le mot-clé d'éveil a été effacé (S610), où le modèle d'éveil par mot-clé fait partie des modèles d'éveil par mot-clé

qui sont associés à différentes informations environnementales, et où la définition du modèle de reconnaissance vocale combiné avec le modèle d'éveil par mot-clé comprend la définition du modèle de reconnaissance vocale combiné au modèle d'éveil par mot-clé qui est associé aux informations environnementales du dispositif, parmi plusieurs modèles d'éveil par mot-clé, en réponse à la réception des informations environnementales du dispositif (S607). La génération du résultat de la reconnaissance vocale inclut l'effacement du mot clé dans le signal vocal reçu, conformément au registre de modèles d'éveil par mot-clé stocké sur le serveur et basé sur les informations environnementales (S609).

12. Un support d'enregistrement permanent compatible avec un ordinateur qui contient un programme qui, lorsqu'il est exécuté sur un système informatique, permet à cet ordinateur d'exécuter la méthode de la

revendication 10.

13. Un support d'enregistrement permanent compatible avec un ordinateur qui contient un programme qui, lorsqu'il est exécuté sur un système informatique, permet à cet ordinateur d'exécuter la méthode de la revendication 11.

5

10

15

20

25

30

35

40

45

50

55

FIG. 1

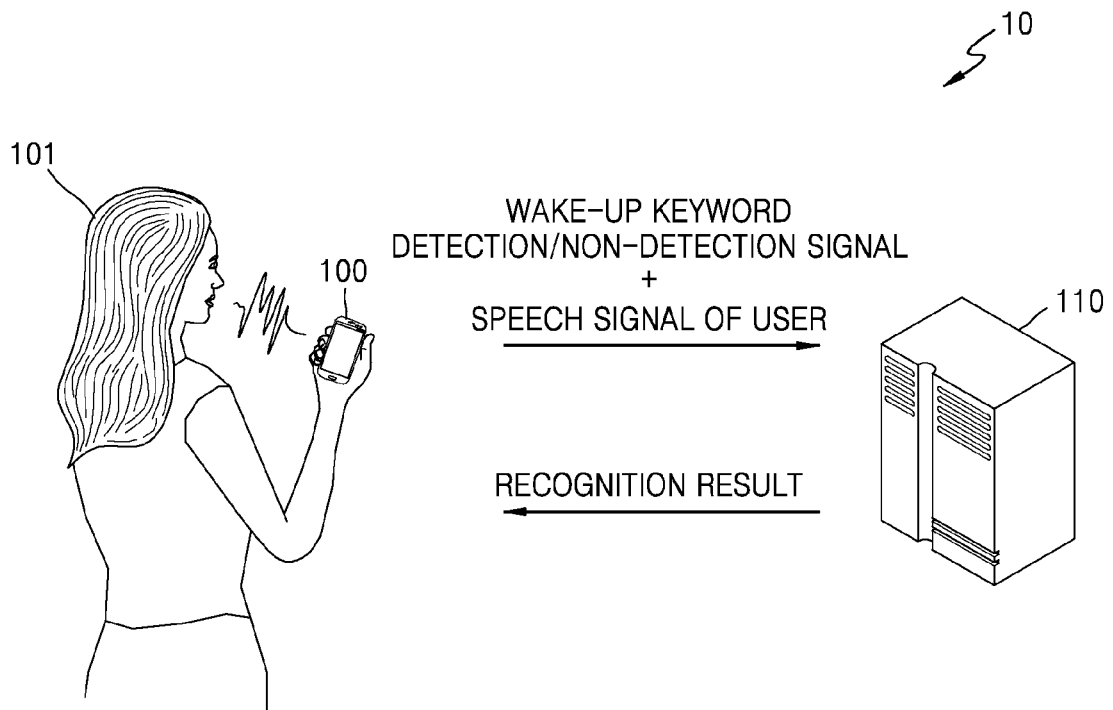


FIG. 2

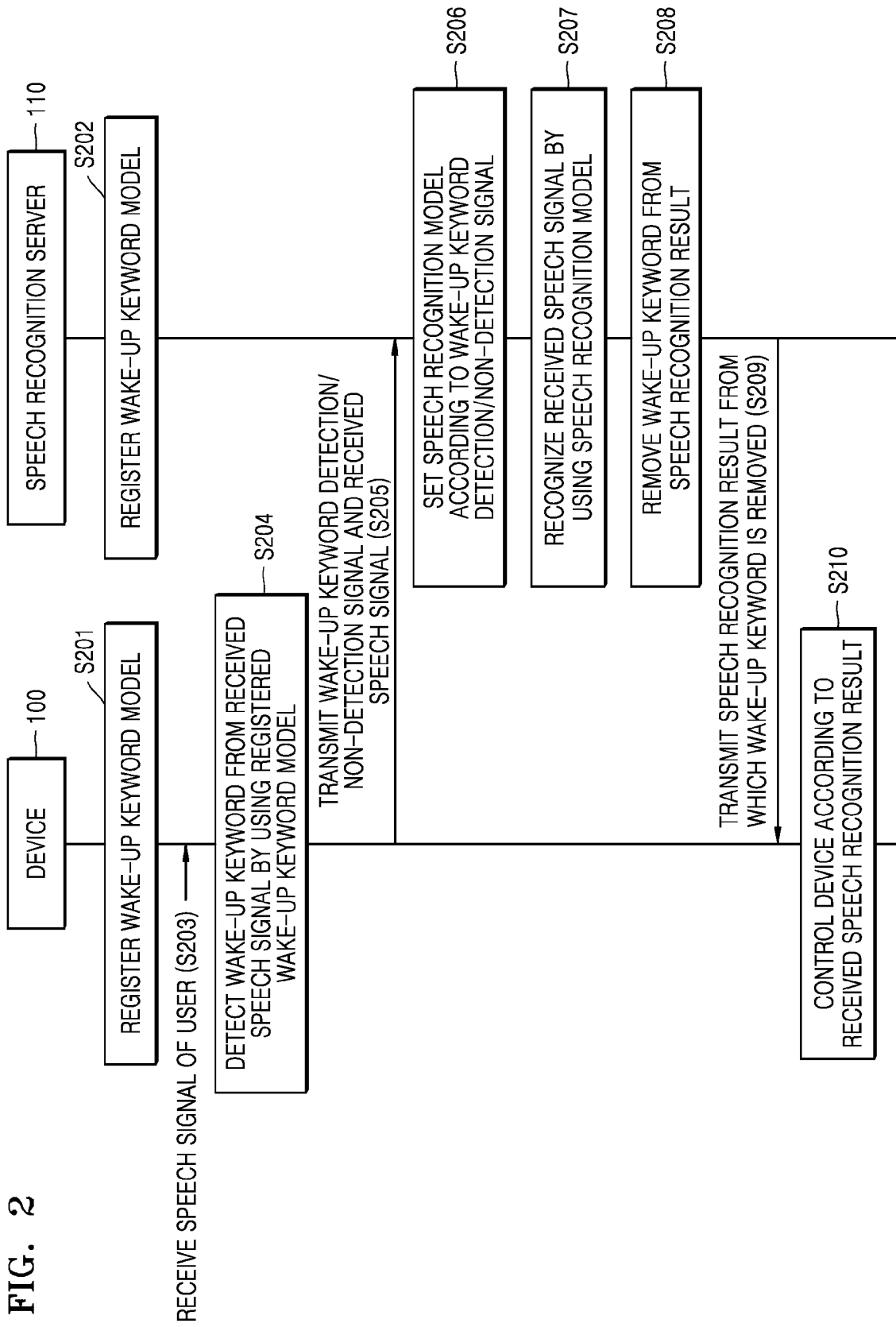


FIG. 3

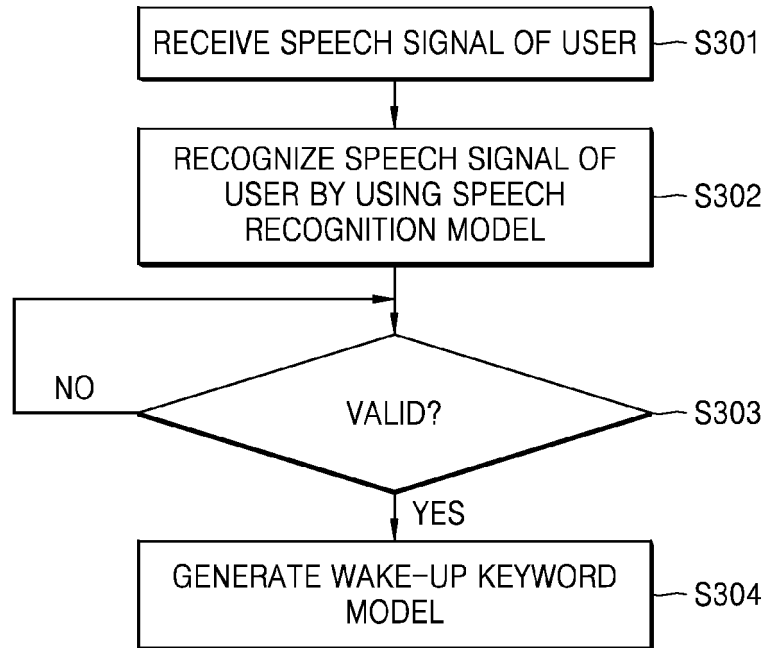


FIG. 4

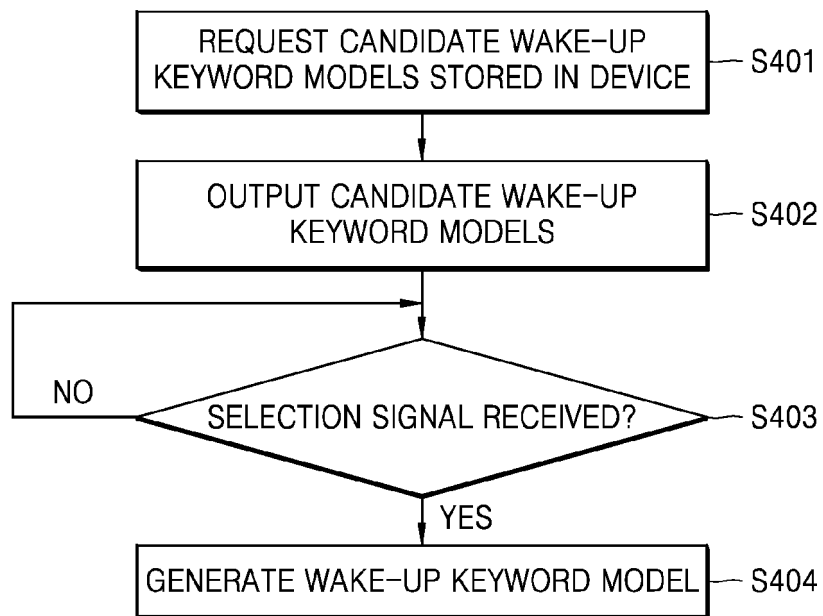


FIG. 5A

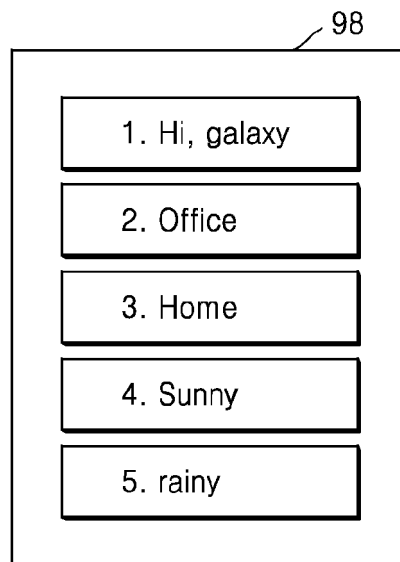
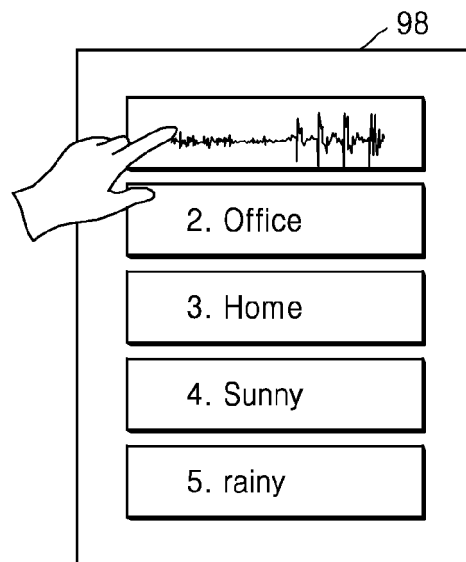
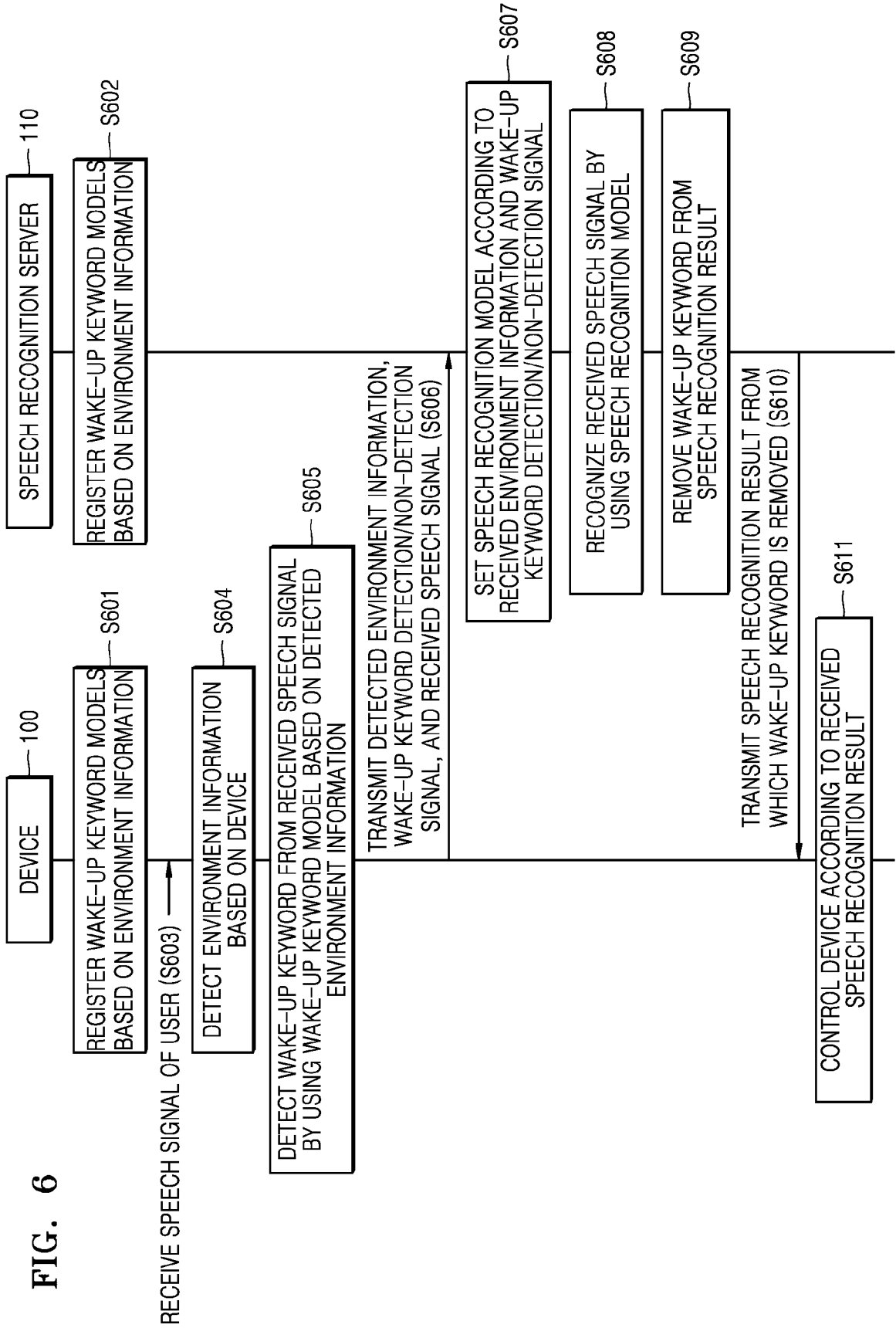


FIG. 5B





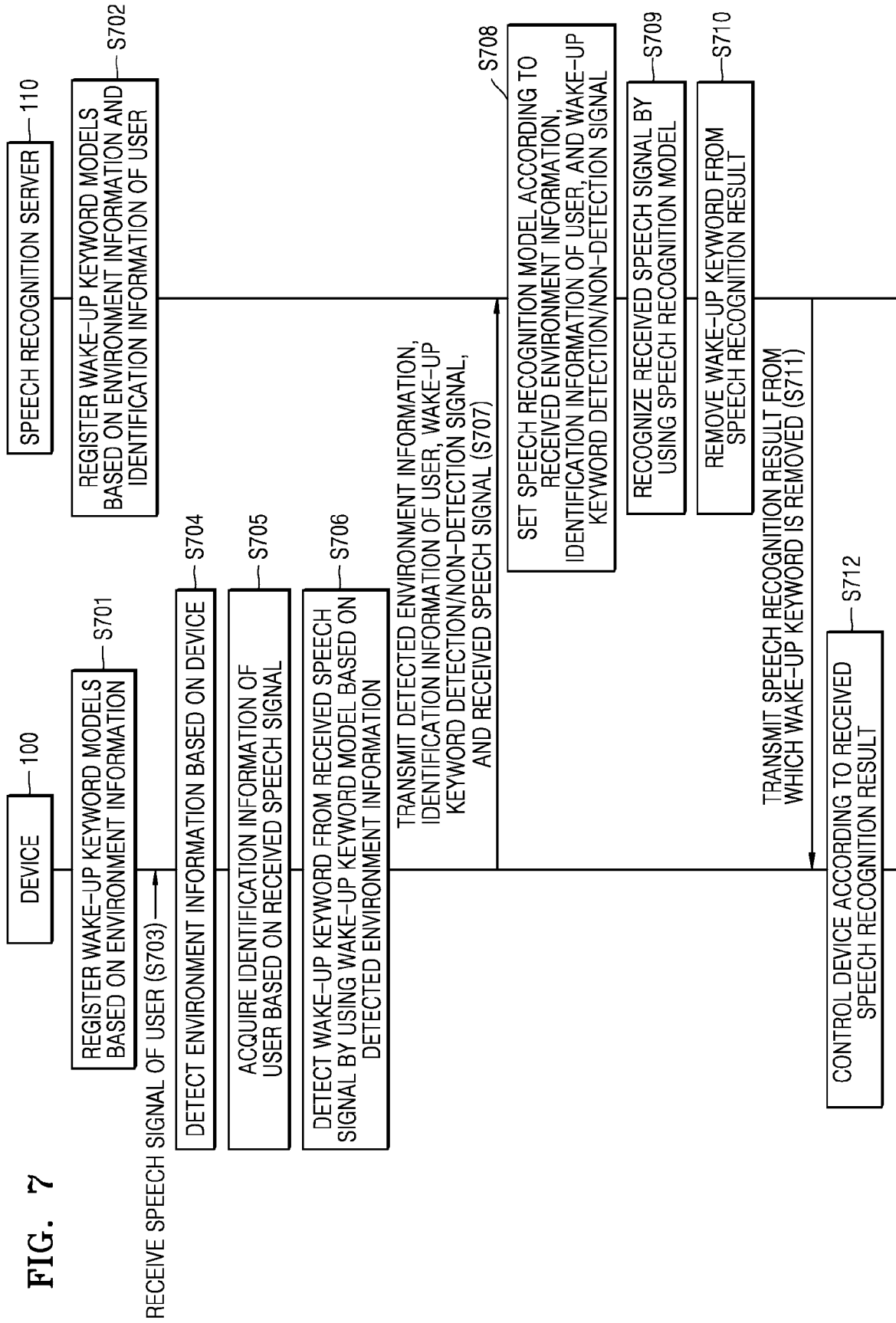


FIG. 8

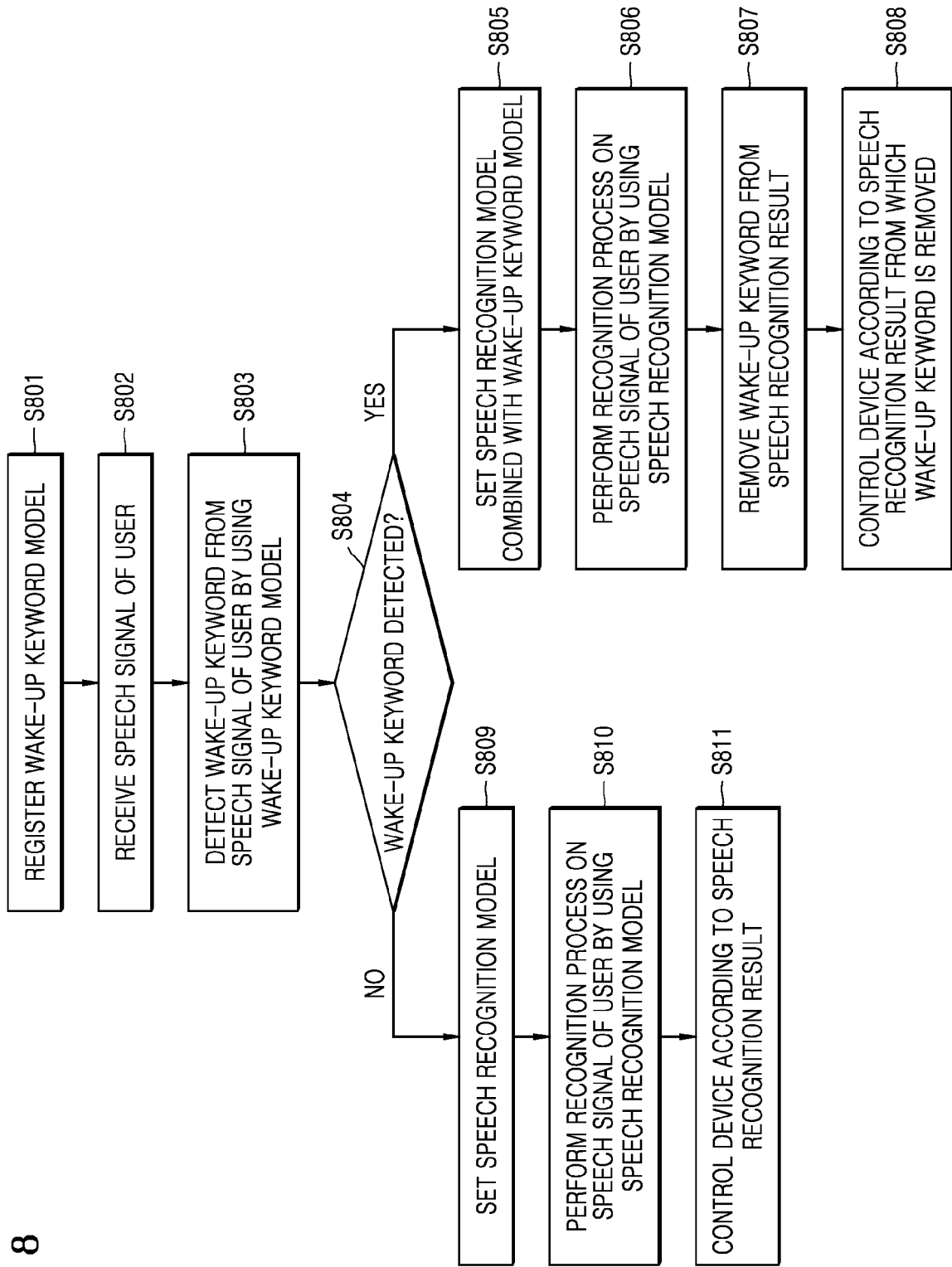
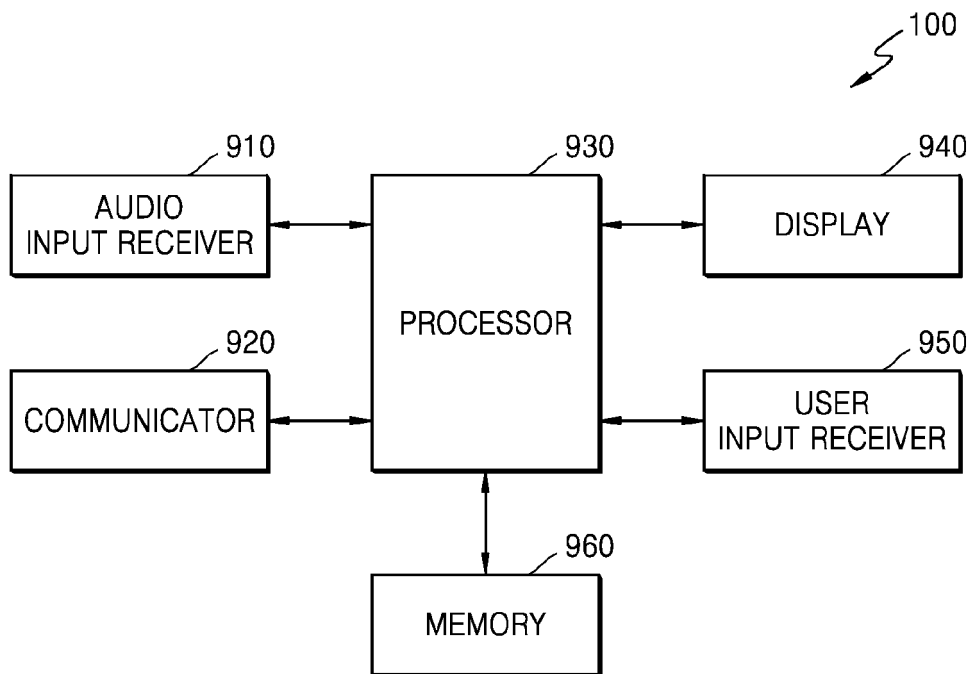


FIG. 9



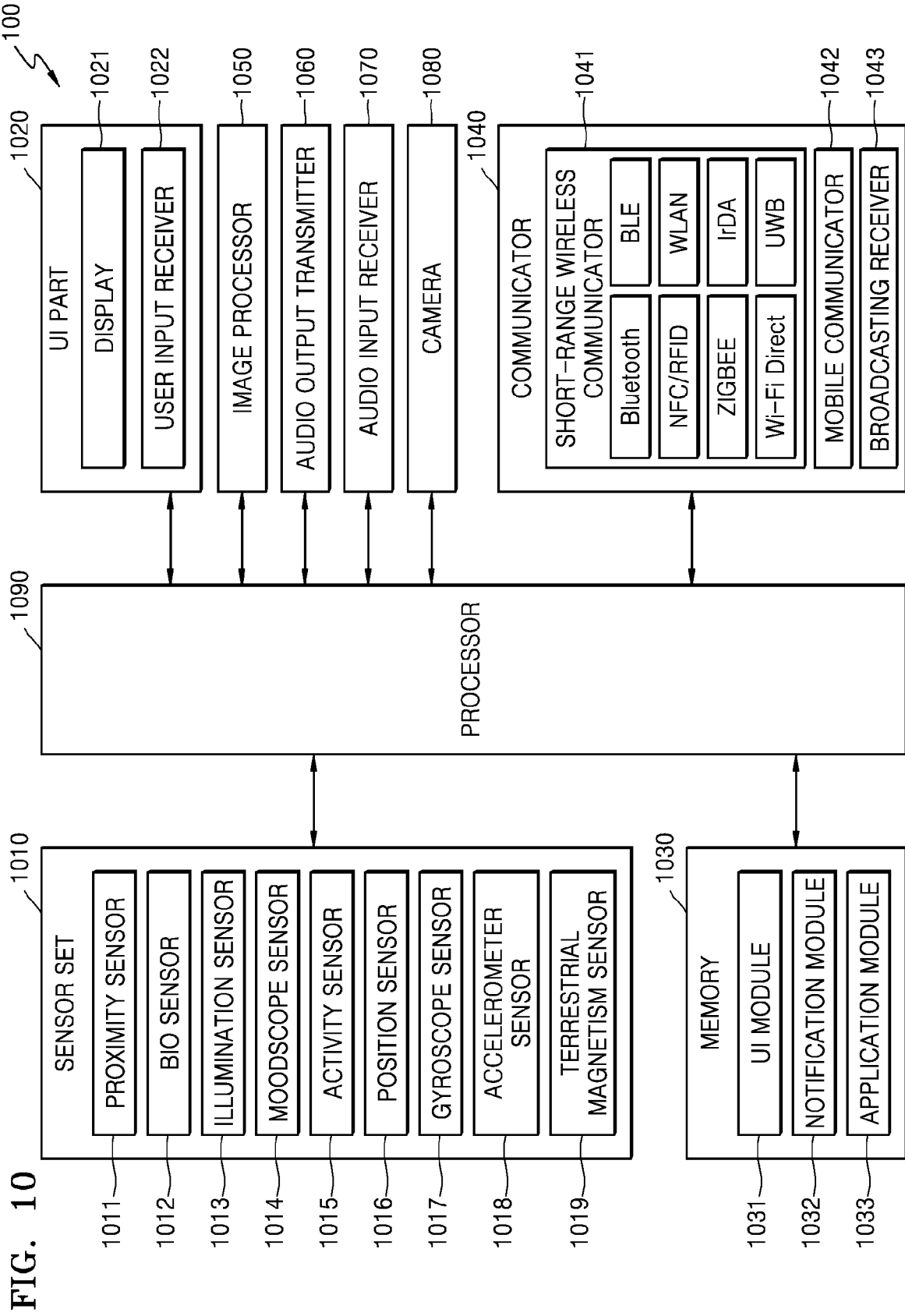


FIG. 11

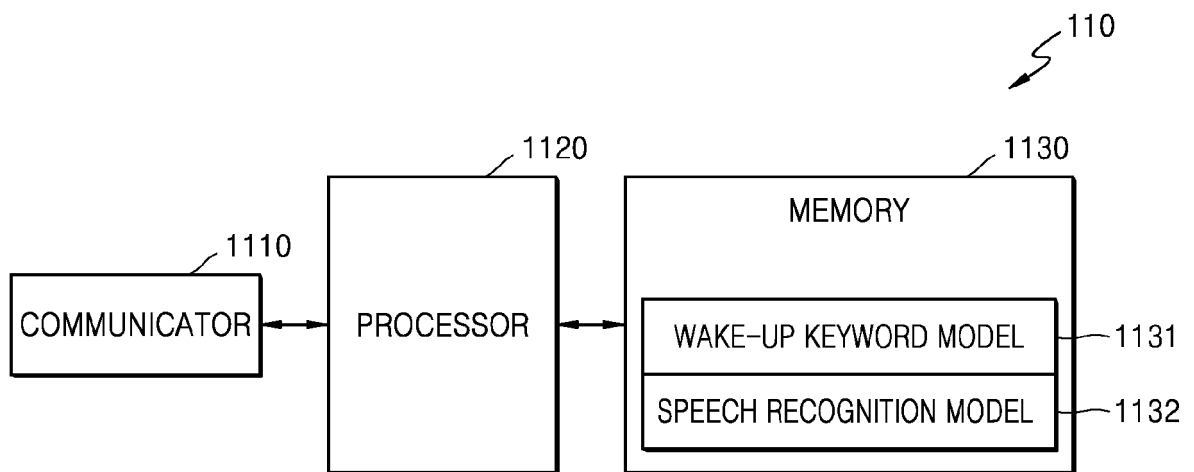
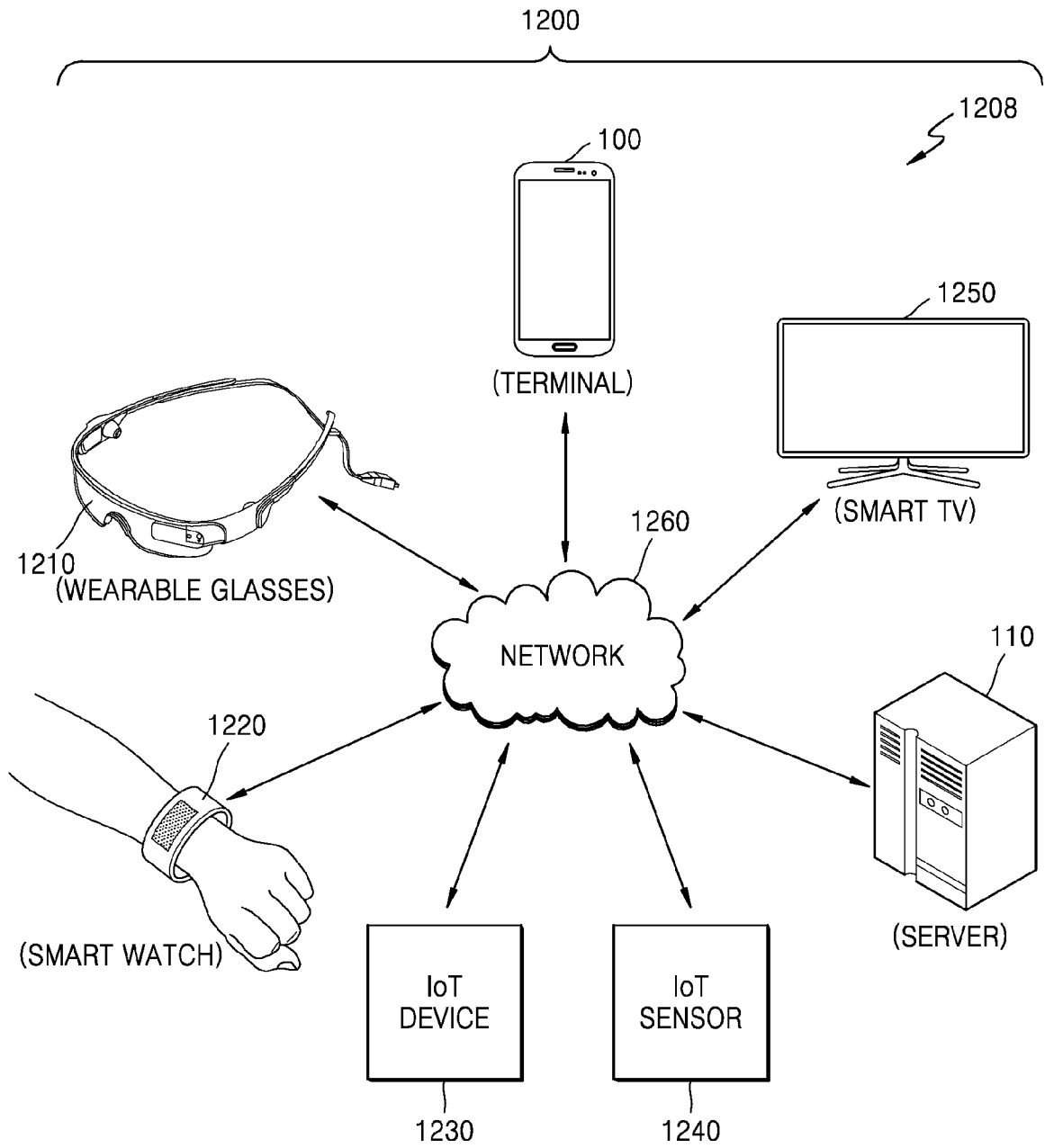


FIG. 12



REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- EP 2669889 A [0006]